

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2006, 22:23:46 ; Search time 146 Seconds
(without alignments)
4897.514 Million cell updates/sec

Title: US-10-067-385-8

Perfect score: 4026

Sequence: 1 KUGTAESKFNGLNGKEGS.....AFVLDKNNISKSTTNNPNK 773

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 92501592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot_7.2.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4026	100.0	2119	2 Q9AHT5_STRPN	Q9Aht5 streptococc
2	4026	100.0	2140	2 Q9TRY6_STRPN	Q97ry6 streptococc
3	3793.5	94.2	2144	2 Q954M8_STRPN	Q954m8 streptococc
4	3709.5	92.1	2144	2 Q8DQ7_STRPN	Q8dqp7 streptococc
5	270.5	6.7	2649	2 Q7RAS7_PLAYO	Q7ras7 plasmodium
6	265.5	6.6	1842	2 Q8IB84_PLAF7	Q8ib84 plasmodium
7	264	6.6	3504	2 Q8IL45_PLAF7	Q8il45 plasmodium
8	260	6.5	1850	2 Q7RGP2_PLAYO	Q7rgp2 plasmodium
9	259	6.4	2757	2 Q7RRR9_PLAYO	Q7rrr9 plasmodium
10	258.5	6.4	1811	2 Q7REH9_PLAYO	Q7reh9 plasmodium
11	257.5	6.4	2661	2 Q7RMS4_PLAYO	Q7rms4 plasmodium
12	257	6.4	1389	2 Q7RPJ4_PLAYO	Q7rpj4 plasmodium
13	255	6.3	1777	2 Q8I3P4_PLAF7	Q8i3p4 plasmodium
14	253.5	6.3	1127	2 Q9YVT6_MSRPV	Q9yvt6 melanoplus
15	252.5	6.3	2227	2 Q8II21_PLAF7	Q8ii21 plasmodium
16	252.5	6.3	2723	2 Q7RQB6_PLAYO	Q7rqb6 plasmodium
17	251.5	6.2	2849	2 Q8IHY4_PLAF7	Q8ihy4 plasmodium
18	251.5	6.2	3381	2 Q8I2V4_PLAF7	Q8i2v4 plasmodium
19	251.5	6.2	3519	2 Q8IE65_PLAF7	Q8ie65 plasmodium
20	250	6.2	1474	2 Q8IIU2_PLAF7	Q8iiu2 plasmodium
21	250	6.2	1650	2 Q77J28_PLAF7	Q77j28 plasmodium
22	247.5	6.1	5767	2 Q8I525_PLAF7	Q8i525 plasmodium
23	246.5	6.1	2033	2 Q8IM18_PLAF7	Q8im18 plasmodium
24	245	6.1	2269	2 Q8ILA2_PLAF7	Q8ila2 plasmodium
25	242.5	6.0	2664	2 Q7RELO_PLAYO	Q7rel0 plasmodium
26	242.5	6.0	3317	2 Q8EWP8_MYCPE	Q8ewp8 mycoplasma
27	241.5	6.0	1033	2 Q8IBB8_PLAF7	Q8ibb8 plasmodium
28	241.5	6.0	2694	2 Q7RJJ1_PLAYO	Q7rjj1 plasmodium
29	241.5	6.0	3063	2 Q6IMC1_9APIC	Q6imc1 plasmodium
30	240.5	6.0	2740	2 Q7RFS2_PLAYO	Q7rfs2 plasmodium
31	240.5	6.0	4433	2 Q8IJ15_PLAF7	Q8ij15 plasmodium

32	240	6.0	1455	2 Q8IKG8_PLAF7	Q8ikg8 plasmodium
33	240	6.0	10061	2 Q8I3Z1_PLAF7	Q8i3z1 plasmodium
34	239.5	5.9	2569	2 Q8IBG8_PLAF7	Q8ibg8 plasmodium
35	239	5.9	2586	2 Q7PDT7_PLAYO	Q7pdt7 plasmodium
36	238	5.9	1081	2 Q8XIL2_CLOPE	Q8xil2 clostridium
37	238	5.9	1339	2 Q4YUUS_PLABE	Q4yuu5 plasmodium
38	238	5.9	3322	2 Q8IKL0_PLAF7	Q8ikl0 plasmodium
39	238	5.9	3628	2 Q968Y0_PLAFA	Q968y0 plasmodium
40	238	5.9	3704	2 Q8IKY8_PLAF7	Q8iky8 plasmodium
41	237.5	5.9	1104	2 Q7RSQ8_PLAYO	Q7rsq8 plasmodium
42	237.5	5.9	2435	2 Q4YTY1_PLABE	Q4yty1 plasmodium
43	237.5	5.9	3535	2 Q8IC29_PLAF7	Q8ic29 plasmodium
44	237	5.9	2511	2 Q8IL44_PLAF7	Q8il44 plasmodium
45	237	5.9	5229	2 Q7RTF4_PLAYO	Q7rtf4 plasmodium

ALIGNMENTS

RESULT 1
Q9AHT5_STRPN PRELIMINARY; PRT; 2119 AA.
AC Q9AHT5;
DT 01-JUN-2001, integrated into UniProtKB/TrEMBL.
DT 01-JUN-2001, sequence version 1.
DT 07-FEB-2006, entry version 21.
DE Serine protease (Fragment).
GN Name=prtA;
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]_SEQUENCE.
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=N4;
RX MEDLINE=21116976; PubMed=11179332;
DOI=10.1128/JAI.69.3.1593-1598.2001;
RA Wizemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,
Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,
Gayle A., Brewah Y.A., Walsh W., Barren P., Lathigra R., Hanson M.,
Langermann S., Johnson S., Koenig S.,
RT "Use of a whole genome approach to identify vaccine molecules
affording protection against Streptococcus pneumoniae infection.";
Infect. Immun. 69:1593-1598(2001).
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
Distributed under the Creative Commons Attribution-NoDerivs License

EMBL; AF291699; AAK19159.1; -; Genomic_DNA.
HSSP; P00782; 2SBT.
MEROPS; S08.064; -.
GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0042802; F:protein self binding; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.
DR GO; GO:0006508; P:proteolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inf_S8A.
DR InterPro; IPR001680; WD40.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF05922; Subtilisin_N; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.

```
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
KW Cell wall; Peptidoglycan-anchor; Protease.
FT NON_TER 1
SQ SEQUENCE 2119 AA; 238228 MW; 517F9B7F6B960A6A CRC64;

Query Match 100.0%; Score 4026; DB 2; Length 2119;
Best Local Similarity 100.0%; Pred. No. 5e-164;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLGEIAESKFNILGNKGESLKDDTTGVEHHQNEESIKESKSFTRDRNISTIRDFENK 60
Db 1313 KLGEIAESKFNILGNKGESLKDDTTGVEHHQNEESIKESKSFTRDRNISTIRDFENK 1372

Qy 61 DLKKLIKKKPREVDDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLEVEYEKLDEIKSKI 120
Db 1373 DLKKLIKKKPREVDDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLEVEYEKLDEIKSKI 1432

Qy 121 YGVLSPSKDGHFELGKISNVSKNAKYVYGNVKSIEIKATKYDFHSHKTMFTFDLIYANIND 180
Db 1433 YGVLSPSKDGHFELGKISNVSKNAKYVYGNVKSIEIKATKYDFHSHKTMFTFDLIYANIND 1492

Qy 181 IVDGLAFAGDMRLFKVNDQKAEIKRMPEKIKETKSEYFVSSYGNVIELGEGDLSKN 240
Db 1493 IVDGLAFAGDMRLFKVNDQKAEIKRMPEKIKETKSEYFVSSYGNVIELGEGDLSKN 1552

Qy 241 KPDNLTKMESKGIYSDEKQOYLKDNIIILKRGYALKVTTYNPGKTDMLGNGVYSKEDI 300
Db 1553 KPDNLTKMESKGIYSDEKQOYLKDNIIILKRGYALKVTTYNPGKTDMLGNGVYSKEDI 1612

Qy 301 AKIOKANPNLRALSETTIYADSRNVEDGRSTQSVLMGALDGFNIIRYQVFTFKNDKGEA 360
Db 1613 AKIOKANPNLRALSETTIYADSRNVEDGRSTQSVLMGALDGFNIIRYQVFTFKNDKGEA 1672

Qy 361 IDKGNLVTDSSKLVLFKGDQKDEYTGEDKNVEALKEDGSMFLFDTKPVNLSMDKNYFNP 420
Db 1673 IDKGNLVTDSSKLVLFKGDQKDEYTGEDKNVEALKEDGSMFLFDTKPVNLSMDKNYFNP 1732

Qy 421 SKSNKIYVRNPEFYLRGKISDKGFNFWELRVNESVDNLYLYGDLHDINTDRDFNKLNVK 480
Db 1733 SKSNKIYVRNPEFYLRGKISDKGFNFWELRVNESVDNLYLYGDLHDINTDRDFNKLNVK 1792

Qy 481 DGDIMDWGMKDYKANGPPPKVTDMDGNVYLOTGYSDLNAKAVGVHYQFLYDNVKEPVNID 540
Db 1793 DGDIMDWGMKDYKANGPPPKVTDMDGNVYLOTGYSDLNAKAVGVHYQFLYDNVKEPVNID 1852

Qy 541 PKGNTSIEYADGKSVFVFNIDKRNNGFDGEIQEQHIYINGKEYTSFNDIKQIIDKTLNIK 600
Db 1853 PKGNTSIEYADGKSVFVFNIDKRNNGFDGEIQEQHIYINGKEYTSFNDIKQIIDKTLNIK 1912

Qy 601 IVVKDFARNTTVKEFILNKOTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGE 660
Db 1913 IVVKDFARNTTVKEFILNKOTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGE 1972

Qy 661 LEKGQFDGWEISGPEGKADAGYVNLISKDTFIKPVFKIIEKKEEENKPTFDVSKKKDN 720
Db 1973 LEKGQFDGWEISGPEGKADAGYVNLISKDTFIKPVFKIIEKKEEENKPTFDVSKKKDN 2032

Qy 721 PQVHNSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISKSSTTNNPNK 773
Db 2033 PQVHNSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISKSSTTNNPNK 2085

RESULT 2
Q97RY6_STRPN PRELIMINARY; PRT; 2140 AA.
AC Q97RY6;
DT 01-OCT-2001, integrated into UniProtKB/TrEMBL.
DT 21-FEB-2006, entry version 22.
DE Serine protease, subtilase family.
GN OrderedLocusNames=SP0641; ORFNames=SP_0641;

OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=AICC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916; DOI=10.1126/science.1061217;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S.N., Heidelberg J.F., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M.L., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Unayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtapple E.K., Khouri H.M., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S.V., Dickinson T.,
RA Hickey E.K., Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
pneumoniae";
RL Science 293:498-506(2001).
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
EMBL; AE005672; AAK74791.1; -; Genomic_DNA.
DR PIR; F95074; F95074.
DR HSP; P00782; ZSBT.
DR MEROPS; S08.064; -.
DR TIGR; SP0641; -.
DR BioCyc; SPN8170187:SP0641-MONOMER; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0042802; F:protein self binding; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0043086; P:negative regulation of enzyme activity; IEA.
DR GO; GO:0006508; P:proteolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR002029; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inf_S8A.
DR InterPro; IPR001680; WD40.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF05922; Peptidase_S8; 1.
DR Pfam; PF00082; Subtilisin_N; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
KW Complete proteome; Protease.
SQ SEQUENCE 2140 AA; 240428 MW; FA44AD8E2938B334 CRC64;

Query Match 100.0%; Score 4026; DB 2; Length 2140;
Best Local Similarity 100.0%; Pred. No. 5.1e-164;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLGEIAESKFNILGNKGESLKDDTTGVEHHQNEESIKESKSFTRDRNISTIRDFENK 60
Db 1334 KLGEIAESKFNILGNKGESLKDDTTGVEHHQNEESIKESKSFTRDRNISTIRDFENK 1393

Qy 61 DLKKLIKKKPREVDDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLEVEYEKLDEIKSKI 120
Db 1394 DLKKLIKKKPREVDDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLEVEYEKLDEIKSKI 1453

Qy 121 YGVLSPSKDGHFELGKISNVSKNAKYVYGNVKSIEIKATKYDFHSHKTMFTFDLIYANIND 180
Db 1454 YGVLSPSKDGHFELGKISNVSKNAKYVYGNVKSIEIKATKYDFHSHKTMFTFDLIYANIND 1513
```

QY 181 IVDGLAFAGDMRLFVKNDQKAEIKIRMPKIKETKSEYYPVSSYGNVIELGEGDLSN 240
 Db 1514 IVDGLAFAGDMRLFVKNDQKAEIKIRMPKIKETKSEYYPVSSYGNVIELGEGDLSN 1573
 QY 241 KPDNLTKMESGKIYSDSEKQVLLKDNIIILRGYALKVTTYNPGKTDMLGNGVYSKEDI 300
 Db 1574 KPDNLTKMESGKIYSDSEKQVLLKDNIIILRGYALKVTTYNPGKTDMLGNGVYSKEDI 1633
 QY 301 AKIQANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFPMNDKGEA 360
 Db 1634 AKIQANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFPMNDKGEA 1693
 QY 361 IDKGNLVTDSSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPVNLSMDKNYFNP 420
 Db 1694 IDKGNLVTDSSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPVNLSMDKNYFNP 1753
 QY 421 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVDNLYIYGDHLHINTDRFNIKLVNK 480
 Db 1754 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVDNLYIYGDHLHINTDRFNIKLVNK 1813
 QY 481 DGDIMDMGMDYKANGFPDKVTDMDGNVYLTQGYSDLNKAVGVHYQFLYDNVKEPVNID 540
 Db 1814 DGDIMDMGMDYKANGFPDKVTDMDGNVYLTQGYSDLNKAVGVHYQFLYDNVKEPVNID 1873
 QY 541 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEQLHIYINGKEYTSFNDIKQIIDKTLNIK 600
 Db 1874 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEQLHIYINGKEYTSFNDIKQIIDKTLNIK 1933
 QY 601 IVVKDFARNTTVKPEILNKDTEVSELKPHRTVTVIQNGKMSSTIVSEEDFILPVYKGE 660
 Db 1934 IVVKDFARNTTVKPEILNKDTEVSELKPHRTVTVIQNGKMSSTIVSEEDFILPVYKGE 1993
 QY 661 LEKGQFDGWEISGPEGKADAGVNLKSDTKPIKPVFKIEBKKEENKPTFDVSKKDN 720
 Db 1994 LEKGQFDGWEISGPEGKADAGVNLKSDTKPIKPVFKIEBKKEENKPTFDVSKKDN 2053
 QY 721 POWNHSQLNESHKEDLQREHSQKSDSTKQVTAIVLDKNNISSSKSTNNPNK 773
 Db 2054 POWNHSQLNESHKEDLQREHSQKSDSTKQVTAIVLDKNNISSSKSTNNPNK 2106
 RESULT 3
 Q9S4W8_STRPN PRELIMINARY; PRT; 2144 AA.
 AC Q9S4W8.
 DT 01-MAY-2000, integrated into UniProtKB/TrEMBL.
 DT 01-MAY-2000, sequence version 1.
 DT 07-FEB-2006, entry version 24.
 DE Cell wall-associated serine proteinase precursor PrtA.
 GN Name=PrtA;
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=3.B;
 RX MEDLINE=21585565; PubMed=11728722;
 RA Bethé G., Nau R., Wellmer A., Hakenbeck R., Reinert R.R., Heinz H.P.,
 RA Zysk G.;
 RT "The cell wall-associated serine proteinase PrtA: a highly conserved
 RT virulence factor of Streptococcus pneumoniae.";
 RL FEMS Microbiol. Lett. 205:99-104(2001).
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 CC EMBL; AF127143; AAD48399.1; -; Genomic_DNA.
 DR HSP; P00782; 2SBT.
 DR MEROPS; S08.064; -.
 DR GO; 0009986; C:cell surface; IEA.
 DR GO; 0005618; C:cell wall; IEA.
 DR GO; 0016020; C:membrane; IEA.

DR GO; 0008233; F:peptidase activity; IEA.
 DR GO; 0042802; F:protein self binding; IEA.
 DR GO; 0004285; F:subtilase activity; IEA.
 DR GO; 0004308; F:negative regulation of enzyme activity; IEA.
 DR GO; 0006508; F:proteolysis; IEA.
 DR InterPro; IPR010435; DUF1034.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR003137; PA.
 DR InterPro; IPR002029; Pept_S8_S83.
 DR InterPro; IPR010259; Prot_inh_S8A.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF06280; DUF1034; 1.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 DR Pfam; PF02225; PA; 1.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR Pfam; PF05922; Subtilisin_N; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR TIGRfam; TIGR01167; LPXTG_anchor; 1.
 DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
 DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
 DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.
 KW Signal.
 FT SIGNAL 1 19 Potential.
 FT CHAIN 20 2144 cell wall-associated serine proteinase
 FT PrtA.
 SQ SEQUENCE 2144 AA; 240726 MW; 2052511470741331 CRC64;
 Query Match 94.2%; Score 3793.5; DB 2; Length 2144;
 Best Local Similarity 94.7%; Pred. No. 4.7e-154;
 Matches 732; Conservative 16; Mismatches 24; Indels 1; Gaps 1;
 QY 1 KLGETAESKFNGLNGKESGLKDDTGYVEHHQNEESIEKSSFTIDRNTSTIRDFENK 60
 Db 1339 KLGEIPESKFNKLVNDDSLNKETAIVENLLVDNQSIKSLFNIHKTTSTIRDFENK 1398
 QY 61 DLKXLIKKKFRVDDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLLEYETEKLDKIKSKI 120
 Db 1399 DLKXLIKKKYQEDDFVN-GGTRTVERYDYKDDKGNIIAYDDGTDLLEYETEKLDKIKSKI 1457
 QY 121 YGVLSPSKDGHFELIGKISNVSKNAKYVYNNYKSIKATKYDFPHSKTMTFDLYANIND 180
 Db 1458 YGVLSPSKDGHFELIGKISNVSKNAKYVYNNYKSIKATKYDFPHSKTMTFDLYANIND 1517
 QY 181 IVDGLAFAGDMRLFVKNDQKAEIKIRMPKIKETKSEYYPVSSYGNVIELGEGDLSN 240
 Db 1518 IVDGLAFAGDMRLFVKNDQKAEIKIRMPKIKETKSEYYPVSSYGNVIELGEGDLSN 1577
 QY 241 KPDNLTKMESGKIYSDSEKQVLLKDNIIILRGYALKVTTYNPGKTDMLGNGVYSKEDI 300
 Db 1578 KPDNLTKMESGKIYSDSEKQVLLKDNIIILRGYALKVTTYNPGKTDMLGNGVYSKEDI 1637
 QY 301 AKIQANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFPMNDKGEA 360
 Db 1638 AKIQANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFPMNDKGEA 1697
 QY 361 IDKGNLVTDSSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPVNLSMDKNYFNP 420
 Db 1698 IDKGNLVTDSSKLVFGKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPVNLSMDKNYFNP 1757
 QY 421 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVDNLYIYGDHLHINTDRFNIKLVNK 480
 Db 1758 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVDNLYIYGDHLHINTDRFNIKLVNK 1817
 QY 481 DGDIMDMGMDYKANGFPDKVTDMDGNVYLTQGYSDLNKAVGVHYQFLYDNVKEPVNID 540
 Db 1818 DGDIMDMGMDYKANGFPDKVTDMDGNVYLTQGYSDLNKAVGVHYQFLYDNVKEPVNID 1877
 QY 541 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEQLHIYINGKEYTSFNDIKQIIDKTLNIK 600
 Db 1878 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEQLHIYINGKEYTSFNDIKQIIDKTLNIK 1937
 QY 601 IVVKDFARNTTVKPEILNKDTEVSELKPHRTVTVIQNGKMSSTIVSEEDFILPVYKGE 660

Db 1938 IVVKDFARNTTVKFIILNKDTGEVSELKPHRITVTIIONGKEMSTIYSEEDFILPVYKGE 1997
Qy 661 LEKGYQFDGWEISGFEGKDKAGYVINLSKDTFKIPVFKKIEEKEENKPTFDVSKKDN 720
Db 1998 LEKGYQFDGWEISGFEGKDKAGYVINLSKDTFKIPVFKKIEEKEENKPTFDVSKKDN 2057
Qy 721 PQVNHSQLNESHKREDLQREHSHQSKSDTKDVTATVLDKNNISSKSTNNPNK 773
Db 2058 PQVNHSQLNESHKREDLQREHSHQSKSDTKDVTATVLDKNNISSKSTNNPNK 2110

RESULT 4
Q8DQP7_STR6
ID Q8DQP7_STR6 PRELIMINARY; PRT; 2144 AA.
AC Q8DQP7;
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2003, sequence version 1.
DT 07-FEB-2006, entry version 19.
DE Cell wall-associated serine proteinase PrtA (EC 3.4.21.-).
GN Name=prtA; OrderedLocNames=spr0561;
OS Streptococcus pneumoniae [strain ATCC BAA-255 / R6].
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX MEDLINE=21429245; PubMed=11544234;
RX DOI=10.1128/JB.183.19.5709-5717.2001;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA Dehoff B.S., Esarem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Kioja H., Kraft A.R., Lagace R.E.,
RA Leblanc D., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenry M., McLeaster K., Mundy C.W., Niclas T.I.,
RA Norris F.H., O'Garra M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.,
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AE008434; AAK99365.1; -; Genomic_DNA.
DR PIR; A97942; A97942.
DR HSSP; P00782; 28BT.
DR MEROPS; S08.064; -.
DR BIOCYC; SPNE1313; SP00561-MONOMER; -.
DR GO; GO:0009986; C:cell surface; IEA.
DR GO; GO:0005618; C:cell wall; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0042802; F:protein self binding; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0043086; F:negative regulation of enzyme activity; IEA.
DR GO; GO:0006508; P:proteolysis; IEA.
DR InterPro; IPR010435; DUF1034.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inh_S8A.
DR Pfam; PF06280; DUF1034; 1.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase S8; 1.
DR Pfam; PF05922; Subtilisin_N; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR TIGRFAMs; TIGR01167; LPXTG anchor; 1.
DR PROSITE; PS0847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN 1.
DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN 1.

DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN 1.
KW Complete proteome.
SQ SEQUENCE 2144 AA; 240438 MW; 8C1B4B1DBC503A0C CRC64;
Query Match 92.1%; Score 3709.5; DB 2; Length 2144;
Best Local Similarity 92.4%; Pred. No. 1.9e-150;
Matches 714; Conservative 26; Mismatches 32; Indels 1; Gaps 1;
Qy 1 KLEIAESKPNLGNKGESLKKDITGVEHHHGHNEESIKEKSFITDRINISTIRDPENK 60
Db 1339 KLGIEISDKLNAKSPENT-NNNAQDESKPDKSSVEGEASLEINKTISTITREFENK 1397
Qy 61 DLAKLKKKPREVDDETSETGKRWEEVDYKDDKGNIIAYDDGTDLVEYETKLEIKSKI 120
Db 1398 DLKKLKKKPREVNDFTSGTKRIIEYDYKDDKGNIIAYDDGSALQYETKFPDEIKSKI 1457
Qy 121 YGVLSPSKDGHFELGKISNVSNKAKVYVYNNYKSIIEIKATKYDFHSHKTMTFDLYANIND 180
Db 1458 YGVLSPSKDGHFELGKISNVSNKAKVYVYNNYKSIIEIKATKYDFHSHKTMTFDLYANIND 1517
Qy 181 IVGGLAPAGDMRLFVKDNDQKAEIKIRMEPEKIKETKSEYYPYSSYGNVIELGEGDLSKN 240
Db 1518 IVDGLAPAGDMRLFVKDNDQKAEIKIRMEPEKIKETKSEYYPYSSYGNVIELGEGDLSKN 1577
Qy 241 KPDNLTKWESGKIYSDSEKQOYLLKONIILKRGVALKVTTYNPGKTDMLSGNGVYSKEDI 300
Db 1578 KPDNLTKWESGKIYSDSEKQOYLLKONIILKRGVALKVTTYNPGKTDMLSGNGVYSKEDI 1637
Qy 301 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFKMNDKGEA 360
Db 1638 AKIQKANPNLRVLSSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFKMNDKGEA 1697
Qy 361 IDKDGNLVTSSKLVLPFGKDDKEYTGEDKFNVAEIKEDGSMLEFIDTPKVNLSMDKNYFNP 420
Db 1698 IDKDGNLVTSSKLVLPFGKDDKEYTGEDKFNVAEIKEDGSMLEFIDTPKVNLSMDKNYFNP 1757
Qy 421 SKSNKIYVRNPEFVLRGKISDKGFWELRVNWSVDNLIYGLDHDNTRDFNIKLVNK 480
Db 1758 SKSNKIYVRNPEFVLRGKISDKGFWELRVNWSVDNLIYGLDHDNTRDFNIKLVNK 1817
Qy 481 DGDIMDGMKDYKANGFPDVTMDGNVYLQTVGSDLNKAKAVGHYQFLYDNVKEPVNID 540
Db 1818 DGDIMDGMKDYKANGFPDVTMDGNVYLQTVGSDLNKAKAVGHYQFLYDNVKEPVNID 1877
Qy 541 PKGNTSIEYADGKSVFVNINDKRNNGPDGIEIQEHYIYNGKEYTSFNDIKQIDTTLNIK 600
Db 1878 PKGNTSIEYADGKSVFVNINDKRNNGPDGIEIQEHYIYNGKEYTSFNDIKQIDTTLNIK 1937
Qy 601 IVVKDFARNTTVKFEILNKDTGEVSELKPHRVTIIONGKEMSTIYSEEDFILPVYKGE 660
Db 1938 IVVKDFARNTTVKFEILNKDTGEVSELKPHRVTIIONGKEMSTIYSEEDFILPVYKGE 1997
Qy 661 LEKGYQFDGWEISGFEGKDKAGYVINLSKDTFKIPVFKKIEEKEENKPTFDVSKKDN 720
Db 1998 LEKGYQFDGWEISGFEGKDKAGYVINLSKDTFKIPVFKKIEEKEENKPTFDVSKKDN 2057
Qy 721 PQVNHSQLNESHKREDLQREHSHQSKSDTKDVTATVLDKNNISSKSTNNPNK 773
Db 2058 PQVNHSQLNESHKREDLQREHSHQSKSDTKDVTATVLDKNNISSKSTNNPNK 2110
RESULT 5
Q7RAS7_PLAYO
ID Q7RAS7_PLAYO PRELIMINARY; PRT; 2649 AA.
AC Q7RAS7;
DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.
DT 15-DEC-2003, sequence version 1.
DT 07-FEB-2006, entry version 13.
DE Hypothetical protein.
GN ORFNames=PY06422;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;


```

RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=17XNL;
RX MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Koop J.T.W., Pertea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shalom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL: AABL01002172; EAA18637.1; -; Genomic DNA.
DR GO: GO:0046872; F:metal ion binding; IEA.
DR GO: GO:0004222; F:metalloendopeptidase activity; IEA.
DR GO: GO:0009405; P:pathogenesis; IEA.
DR GO: GO:0006508; P:proteolysis; IEA.
DR InterPro: IPR011591; Botulinum.
DR Prodom: PD001963; Botulinum; 2.
DR Complete proteome; Hypothetical protein.
KW SEQUENCE 2649 AA; 309812 MW; E7207F344643AC24 CRC64;

```

```

Query Match      6.7%; Score 270.5; DB 2; Length 2649;
Best Local Similarity 20.6%; Pred. No. 0.0064;
Matches 206; Conservative 160; Mismatches 327; Indels 307; Gaps 49;

QY 1 KLGEIAESKFKNLGN-----GKESLKKDTTGVHHHQNEESIEKESFTIDRNISTRD 56
Db 1044 KNNLIERNYRNDNLHVNKAKKNNKDISLNLYNSKSTGDNVQVRSVITETETIRS 1103
QY 57 FE-----NKDLKKLKKKFRVDDFTSTGKRMEEYDYKDKGNIAYDD----- 102
Db 1104 FEDKIKLKLQNERLKKKIEKLYDEKVKQNDY-KMEKIKKQD---NLFEATDKHIEK 1159
QY 103 -GTDLEYETEKLDKSKIKYGLVSPKDGHEILGKISNVSKNAKYVGN--NYKSIEIK 159
Db 1160 LHCELENKSKQNELIKINL-----KDKMKIELESQICNN-NVHSNEHLNHTLDVK 1211
QY 160 ATKYDFHSKWTFFDLYANINDIVDGLAFAGDM-----RLFVKDNDOKAE-IKIRMPEKI 213
Db 1212 NSENNKTEHTNNNDVHLSNDSITDKKIKYQVILLQDQFLMKIEIKQWFLSKNVELN 1271
QY 214 KETKSEYYPVSSYGNVIBELGEDLSKNKPDNL----- 245
Db 1272 KLANMKHINNEYERSIDKLEKSIDLKEQNLKLTQICDLKEKNIMLEKAAQLNDESS 1331
QY 246 ----TKMESG-----KIYSD----- 256
Db 1332 NTTISSDGTGINNEIKIMKEIEBALYKDKIKLLSKNLEETKNKINTLLTKTSNBQS 1391
QY 257 ---SEKQOYLKNDLILKRGVALKVTYTPNGKTDMLCGNVVYSKEDIKAIQKAMPNLRAL 313
Db 1392 IELNKKIKCLLKNKLNQKVEKSI-----NDLKKLNKYD-EDILKENSIPKDSII- 1442
QY 314 SETTIYASRNVEDGRSTQSVLMSALD---GFNIIR--YQVFTFKQNDKG-----EAD 362
Db 1443 --TFDLNNEKTEHTNVKQDTESTQIDSDSDYDNWRIKKGVDISNHNKDNSENCVQDKTEID 1500
QY 363 KD-----GNL---VTSSKLVLFEGDKDEY--GEDKFNVEAIKEDGSMFLDTDKV- 409
Db 1501 KMVEKCTPNLKETQTDITNLVLISVDKEMTEKNQDDYN---ILEDNNLIKINETSMF 1557

```

```

RESULT 6
ID Q8IB84_PLAF7 PRELIMINARY; PRT; 1642 AA.
AC Q8IB84;
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2003, sequence version 1.
DT 07-FEB-2006, entry version 7.
DE Hypothetical protein MAL8P1.29.
GN Name=MAL8P1.29;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,
RA Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL: AL844507; CAD51123.1; -; Genomic DNA.
KW Hypothetical protein.
SQ SEQUENCE 1642 AA; 194754 MW; 701F5D28B2BE8097 CRC64;

Query Match      6.6%; Score 265.5; DB 2; Length 1642;
Best Local Similarity 20.4%; Pred. No. 0.0059;
Matches 190; Conservative 133; Mismatches 313; Indels 297; Gaps 46;

QY 4 EIAESKFKNLGNEGSLKKDTTGVHHHQNEESIEKESFTIDRNISTIRDENK 60
Db 104 EINERNNNNNNNNNNSNSFIMDEKEKTNNTMLLKHKKRYDPPDNTS-----DEK 158
QY 61 DLKLLKKKFRVDDFTSETGKRM-----EYDYKDDKGNIIAYDDGTDLEYETEKLDE 115
Db 159 NIKR-----KKINDIINKYKKMIIPSYENDSKYSGSFNISDD-----DE 201
QY 116 IKSIIYGLVSPKDGHEILGKISNVSKNAKYVGNNYKSIIEIKATKYDFHSKWTDPOLY 175

```



```
Db 2106 DNKKKFDNVLEIPKIGHNI-----LDDKETITEQVEEKSIGQDKSMENNNVST--NDG 2157
QY 640 KEMSSITVSEEPFILPVYKGELEKGYQFDGWEISGEGKKDAGYVNLK-----689
Db 2158 KOIH---IQEDI-----KXNIIINVNDKHSKSNLNLHIDEPKHYVEEKIKKH 2204
QY 690 ----DTIKPVFKKIEBKKEENKPT-----FVSKKKDKNPQVNHSQLNSHREKDLQREE 741
Db 2205 ETADHDIKKEFKIEQDENSQNEPSNENILVDVNAQDDK-----NISKLNDLHDQE 2256
QY 742 HSKQSDSTKDVATVLDLKKNISKS 766
Db 2257 KGTNDSVVE-----HNVSOKT 2273

RESULT 8
Q7RGP2 PLAYO
ID Q7RGP2_PLAYO PRELIMINARY; PRT; 1850 AA.
AC Q7RGP2;
DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.
DT 15-DEC-2003, sequence version 1.
DT 07-FEB-2006, entry version 13.
DE Hypothetical protein.
GN ORFNames=PY04304;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=17XNL;
RX MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,
RA Silva J.C., Emolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feidlyum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabli A., Cummings L.M.,
RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii."
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
CC EMBL; AABL01001295; EAA16146.1; -: Genomic_DNA.
DR Complete proteome; Hypothetical protein.
KW SEQUENCE 1850 AA; 215317 MW; 00366A8D5CE21629 CRC64;

Query Match 6.5%; Score 260; DB 2; Length 1850;
Best Local Similarity 19.2%; Pred. No. 0.012;
Matches 198; Conservative 154; Mismatches 379; Indels 298; Gaps 44;

QY 6 AESKFNILNG-----KGSLLKOTGVHHQNEEESIKEKSFTRDNRISTIR-----55
Db 62 SEEKNDQNGKGYKFNNSSTNKKSTNDSSYDNLADYVINE-----DNNLSFNFSVCIN 115
QY 56 ---DPEN-KOLKLLKKKF--REVDDFTSETGKRMEEYDK---YDDKGNIIAYD---D 102
Db 116 SKLDYSNILNRKSLBKPNHNNDDSDNSDEYDLYGFRNIRRYKECSFNADKTKD 175
QY 103 GTDLEYETEKLDEIKSIYGVLSLP-----SKDGHFEI-----134
Db 176 NAEKYSLEKYNTIDN--IGTIAPINSIGDNLKRGIKTCTNISEYNSMNNKIEVNATS 233
QY 135 ---LGKISNVSNKAVYVGNNGYKSIIEIKATKYDFHSKTWTF-----172
```

```
Db 234 LENVKYNENDKYNMEYKNDNDYVKINTKDSINQREKNIYKSPSTDAILSRIKSDNDN 293
QY 173 -----DLYANINDIVDGLAFAGDM-RLFPKNDQKKAIEK-----206
Db 294 LCDNKDDIYANMTDLNKMVMQDGMKKLYTDGNINKTPTKTFKSYFYLPKPTAGKN 353
QY 207 ---IRMPKIKETKSEY-----PYVSSY-----GNVIELGEGDLKSNKPNLNTKMESGKIY 254
Db 354 ISELETPNLKDDKEDYSGNTFFIAYILAGQPENKSLNGEENLENKISFPNK-EIDKDY 412
QY 255 SDSEKQQYLLKDNIIILRG-----YALKVTYTPNGKTDML--EONGVYSKEDIAK 302
Db 413 SKENFHEPFFKFDINTKEGDEPEBEABRTEIEENTFNIIKDNFSKKNHSPSNNI 472
QY 303 IQKANPN-----LRALSETT-----IYADSRNVED-----CRSTQSVLMSAL 339
Db 473 VNQMNPNDNKDELDIFRRVSVTSKKYNFIKTEEYNDMLKERDIKLSSNKSVSRSKTNA 532
QY 340 DGFNIIRYQVFTFK-----MNDKGEAIDKGNLVDSSKLVLFQKDDKEYTGEDKFN 391
Db 533 NGLDILE-KVNTKKNINNDVSFTDKIFSCNNFNIMNTTTPKDNLF--KEPKYSKLDHNE 589
QY 392 VEAIKEDGSMFLI-----DTXPVNLSDMKYFNFSKSNKIYVRNPEFYLGRKISDK 442
Db 590 INFERYKGMLLYSINDKNGINENENIGLSKNITFTNTNEDNYDQIDRTIPLNKLRLSM 649
QY 443 GGFNWLVRNVSFV---DNYLIYGLDLHDNTRDNFKLVNKGDDIM--DWGMKDYKANG 496
Db 650 YKSYDDIINEGNINNNNDNKNIYNPNLTGSMKF-----IPNSSVMINQWQFKQDINN 704
QY 497 FFDKVTMDGMGYLOTGYSDLNAAVGVHYOFLYDNVKEPVNIDPKGNTSIEYADGKSV 556
Db 705 IPINRSKYTNLIIGNSNSDIYKNTNNIN-----NNKFMNSNDFKSGVQIYVDTEGENIN 759
QY 557 FN-----INDKRNNG-----FDGEIOEQHIYINGKEYT 584
Db 760 FRNSYESKIRNIEKYINDKNNAIRKSNYINKYMTTHDNTFCGDT-----INKMNTN 813
QY 585 SPN-DIKOIIDKTLNIKIV-----VKDFARNTTVKEF-----ILNKDTGEVSELK 628
Db 814 NFHDNPMLLINNRINMGNSVKGDNINKVENEQNTSEQFFKQNDITNYSNGNGD-N 872
QY 629 PHRVTVTIQNGKEMSSITVSEEDFILPVYK---GELEKGYQPDGWEISGF--EGKDA-- 681
Db 873 NSKVNLMEKSNFTFKSIISENDGPIPLNKYIANRSLSETADP-GYKESPYMEESKDEHN 931
QY 682 -----GYVNLKSD---TPIKPVFKKIEBKKEENKPTFDVSKK 717
Db 932 KNFGIHTLGNNDINNENNNNGYNHDIITKYVWDFPKKSFNII-DVNEIKKNILNDYAKV 990
QY 718 KDNQVNHSQLN-----ESHRKEDIQREHSHSQSDSTKDVATVTL-----D 758
Db 991 NNNSENKANMDMEKNSFYKDPNEDTLAKYIHOIKGSI TEVVEKILSHKEINFEINGE 1050
QY 759 KNNISKSST 767
Db 1051 KNNIIKKT 1059

RESULT 9
Q7RRR9 PLAYO
ID Q7RRR9_PLAYO PRELIMINARY; PRT; 2757 AA.
AC Q7RRR9;
DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.
DT 15-DEC-2003, sequence version 1.
DT 07-FEB-2006, entry version 12.
DE Rhoptry protein.
GN ORFNames=PY00649;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
```


Matches 179; Conservative 117; Mismatches 312; Indels 215; Gaps 40;

QY 11 KNLGKGEGLSKDTTGVHHHQEENES-----IKEYSSFTIDRNIISTIRDFENKO 61
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 555 KNINNSHGSIENVNRIIHNNNDKNVSNNHILKNDETKVIDNVDNANTYTGF 614

QY 62 LKKLKKKPREVDFTSETGRMEBYDYKDKGNIIAYDGGTDLEY-----EFEKUDE- 115
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 615 ILS-KDEDPKLETFLNLDKWGSFKLFPEKWLDCSKWQIDLKLOREYINNIELEKKEE 673

QY 116 -----IKSKIYGVLSPKDGHEPILGLKSIVSKN-AKVYGYNNYSIEIKATKYDFH 166
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 674 KKYLEKICLNKI KELWCGRVAGLMNV--KLCNMEKIANEFLSNNIIKNNIHEVTUDYE 731

QY 167 SKTMTF-DLYANINDIVDGLAFAGDMRLFVKNDOKAEIKIRPEKIKETKSEVPYVS 225
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 732 RSAMCYQDLIA-----QLYVICLSK-----KLEMKKINEFK----- 764

QY 226 YGNVIELBGDSLUSKNPNLTQMBSGK-----IYSDSEKQQVLLKNLILRKGYAL- 276
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 765 --NNID-GVANSQENESDAKKKDSKKKGYYVHSLIYSFNKSH---KQN---KKKYSIN 815

QY 277 ---KVTYNPGKTMLNGVGVYKEDI AKIQCANPLRALSETTIYADSRNVEGRSTQS 333
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 816 NDTQT-T-EPSKKD-----NLAKAENERTATTANNNEEDRKIEIN 853

QY 334 VLMSALDGNRIRYQVTFK-MNDX-----GEAIDKOGNLVTDSSKLVLFGKDD--KEYT 385
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 854 KQKG-----NNIENTVNTKNLNEQYKRNSEDYTKRSNIPTDXHNPNFGKRNSYNEKT 908

QY 386 GBDKFNVFAIKEDGSMLEFDTPKVNLSMDKNFYNP-SKSKIYVRNPEF-----YLROKIS 440
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 909 VEDK-----EIKETNKSINVEYYSNNSYENNMSYLDNSSFKNKLNKSEIV 953

QY 441 DKGGFNWELRVNESVVDN--YLIYGDHLIDTRDPIKLVKDGDIMDWGMKDYKANGFP 498
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 954 EKSNIN-----IKKKVDVGGSGFFGSF-----FMFK-----SDYNKKG-- 987

QY 499 DKVTDMDGNVILQTYGSLNKAQAVGHVQFLYDINVKPEVNIIDPKGNTSIEYADGKSVVFN 558
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 998 SQIESNNSAIVEPKSDDSNE---GSDYSFKYNNKEDCKRMKKNSKSNWSAFKK-IFN 1043

QY 559 INDKRNNGDGEIQ---BOHIYINGKEYTSFNDIKQIIDTKLNIKVIYVQFARNTTVKEF 615
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 1044 LSRKKKKNQGDKPSLKSDEKIKKDKNTDINSVENNEDK-----YNNESSNQ 1091

QY 616 ILNKDTGEVSELKPRVTTVITONG-----KEMASTIVSEBDFILPVYKGELEGYQFDG 669
:||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 1092 SSNEEDGHINNYKKO-----NGMKKANRGNGNLNNKSSLESLEBFKFEDEFQFI-- 1142

QY 670 WEISGFEKGDAQGVYINLSKDTFIKVPFKTIEEKEEBENKPTFDVSKKKNQPNVHSOLN 729
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 1143 -----LKKEVGDLHIEDDVESKIIPERFDEDE-----PDIGNKSNSSCSNS--N 1187

QY 730 ESHRKEDLQRBEHSGKSDSTKDVDTAT-VLDKNNISSTKSTNNP 771
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 1188 DSNDSSD-----SNKSIRQSDGTGTQITKTKSSSNKSEKKKP 1224

RESULT 11

Q7RMS4 PLAYO PRELIMINARY; PRT; 2661 AA.

ID Q7RMS4_PLAYO AC Q7RMS4_ DT 15-DEC-2003, integrated into UniProtKB/TREMBL.
AC Q7RMS4_ DT 15-DEC-2003, sequence version 1.
DT 07-FEB-2006, entry version 13.
DE Rhoptyr protein (Fragment).
GN ORFNames=PY02104;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxId=73239;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RC STRAIN=17XNL;
RX MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Portea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shalom S.J., van Aken S.E., Riedmuller S.B., Feldblum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -!- CAMEL: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; AABL01000575; EAA21526.1; -; Genomic_DNA.
DR InterPro; IPR011561; Pox II rel.
DR ProDom; IPR006499; Reticulocyte bd.
DR TIGRFAMs; TIGR01612; 235kDa-fam; 1.
KW Complete proteome.
FT NON TER 2661
SQ SEQUENCE 2661 AA; 312273 MW; 0BD02AD0B712253B CRC64;

Query Match 6.4%; Score 257.5; DB 2; Length 2661;
Best Local Similarity 20.0%; Pred. No. 0.023;
Matches 179; Conservative 151; Mismatches 294; Indels 271; Gaps 41;

QY 22 KQTTGVHHQHNESIK-EKSFITDRNIS-----TIRFENK-- 60
DB 427 KKKLETTIKLKNENEETVTKLEIRELFKYSDEVAEKGVHEELKJLKETIKDIYNKKE 486
QY 61 -----DLKKLKKKFFREVDFTSETGRMEEDYKYDD-----KGNLIAYD 101
DB 487 YIKAIIDLKAIENNNKIYDELGNTPFQIEYVKKDITYSTIKSELSEIYKGNIVELY 546
QY 102 DGTDLLEYETELDKISK-----IYGLSPSKDGHFELGKISNVKN----- 144
DB 547 NELSSVVQENTIDIPKKNKELETLKSEIDNVYKNKIQNMKLEEVLEHLKNTIETSKNELSNL 606
QY 145 ----AKVYGG--NNYKSIEIKATKYDFHSKTMTPFDLYANINDIVDGLAFAGDMRLFVKDN 198
DB 607 ILEIITKIVFGEIDN---DLNKLTKLPKPNKEQ--ELSNKIND-----YTKEN 647
QY 199 DQ-----KAAEIKIRMEKI-----KETSEYPPVSSYGNVIELGEGDLSK----- 239
DB 648 DQLSVYOSKIIETIRKKHYNDQINDVNTKGEAKQNYDOSKEYAKKIPIKEGETSKFINEVT 707
QY 240 -NKPNDNLTKME-----SGKIYSDSEKQVLLKDNII-----LRKGVALKVTTYNPG 284
DB 708 AMKNELRLKKYIDFGAYEKEXANSEHEKITELVKKIRIEISKYLVNVEKK---FNNS 764
QY 285 KTDMLEGNQVYSKE--DIAKIQKANPNLRALSETTIYADSRNVEDGRSTOSVLSALD-G 341
DB 765 KKLIIETDKSIEEYKYNITLKKVDEYIKVCVHT-----NELISFNPNKQITLKDLDKN 819
QY 342 FNIRY-----QVFTFK-----MDKGAIDKGNLV-----TDSSKLVLFP-----G 378
DB 820 INTVKENNSDKLTKYTNKFENILTDKKTLELDTFTDVALNDHESNNMLMKYFNNLKANLG 879
QY 379 KDDK---EYTGEDKFNVEAIKEDGSMLEFDTKTPVNLSDMKNYFNPSPSKNKIYVRNPEFY 434
DB 880 KKKQWMLYQDFGEKEKAIIDIKKNADINENVSKIEVAYIYASINISDEWERIEKSIES 939
QY 435 LRGISDKGGFNWELRNVESVDNYLIYDGLHDINTDRFNIKLNKVKDGDIMDWGMKDYKA 494

Db 940 LNTQVIEK-----VKTNTVNLNEIKGLKQYDFHDFGKEENIK----- 977
Qy 495 NGFPDPKVTMDGNYVLOTGYSDLNAKAVGHYQFLYDNVPEVNDPKGNSTSEYADGKS 554
Db 978 --YDEINKNEI-----KTVG--QQIDQHINKLEBEIKKSGSXNDEIKGKT 1021
Qy 555 VVP-NINDKRNNGPDGIBQOHIYINGKEVTSFN-DIKQIDKTLNLIKIVVDFARTTV 612
Db 1022 DKFENVTD-----TIYNKDPKEIEEIKNI-VTKIDEKQNIYI 1059
Qy 613 KEFTLNKDTGEVSELKPHRVT-----VTIQNGKMSSTIVSEDFILPVYKGELEKGYQ 666
Db 1060 K---INKLLEISEIKKNYLSKKVDINVSYGSLKLFLEQID----- 1101
Qy 667 FDGWEISGFGKDGAGVYVNLKSDTFIKPV--FKKIEEKEEENKPTFD-----VSKKDN 720
Db 1102 -----BEKKAGHMIK-SMEBYMDLDNKKKQSEIENEMKINNDIKMVKEMEA 1151
Qy 721 PQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKKN---ISSKSTNNPN 772
Db 1152 LNTSHDDDKQVH---NISKKEEKISD-----IHKNSLKTIOBSFTESNIN 1194

RESULT 12

ID Q7RPJ4_PLAYO PRELIMINARY; PRT; 1389 AA.
AC Q7RPJ4;
DT 15-DEC-2003, integrated into UniProtKB/TrEMBL.
DT 15-DEC-2003, sequence version 1.
DT 07-FEB-2006, entry version 12.
DE Hypothetical protein.
GN ORFNames=PY01465;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_TaxID=73239;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=17XNL;
RX MEDLINE=22255706; PubMed=12368865; DOI=10.1038/nature01099;
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Portea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
RA Shallow J.J., van Aken S.E., Riedmuller S.B., Feidblyum T.V.,
RA Cho J.K., Quackenbush J., Sedgah M., Shoabi A., Cummings L.M.,
RA Florens L., Yates J.R. III, Raine J.D., Sinden R.E., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;
RT "Genome sequence and comparative analysis of the model rodent malaria
RT parasite Plasmodium yoelii yoelii.";
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DDAJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NonCommercial License
CC -----
CC EMBL; AABL0100391; EAA20802.1; -; Genomic_DNA.
DR GO; GO:0046872; F:metal ion binding; IEA.
DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.
DR GO; GO:0009405; P:pathogenesis; IEA.
DR GO; GO:0006508; P:proteolysis; IEA.
DR InterPro; IPR011591; Botulinum.
DR ProDom; PD001963; Botulinum; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 1389 AA; 164817 MW; 21851CA630F5D9F CRC64;

Query Match 6.4%; Score 257; DB 2; Length 1389;
Best Local Similarity 20.4%; Pred. NO. 0.011;
Matches 193; Conservative 139; Mismatches 288; Indels 326; Gaps 48;

Qy 17 KEGSLKQDTTGVHHQOE-----NEESIKEKSSFTIDRNI STIRD 56
Db 354 KKEAFQKEKNSSEHLHEKNIFYEKEILEASKYDTVSYFNKEIKKOKSSYDNNKGWIPS 413
Qy 57 FENKDLKLLIKKIFREVDFTSETGKRMBEYDYKDDKGNIIAYDDGDTDLEYETEKLEDEI 116
Db 414 EINKQORINKCNKHFVSHSDT---EKVEY-----SNIIS-----DLYNQKNDK 457
Qy 117 KSKIYGVLSKDGHEFELGKISNVSKNAKVYGN-NYKSEIKAT-----KY 163
Db 458 KLIYGL---SKEDY-----SNDKKNQDTEEVYDNDKTDIDQVNDKNVLDISPLTHKY 510
Qy 164 DFHSK-TMTPDLVANINDIVDGLAFAGDMELFVKDNDQKABIKIR---MPEKIKETKSE 219
Db 511 KYRSNLTINKINKIEKII-----EINLNTDVLQKIPGMRIENVILPHK----- 556
Qy 220 YPVSSYGNVI-----ELGEGELSKNKPDLTKMESGKIYSDSEK-----QOYL 263
Db 557 -----YSNDILAKIQIKTLSEENELIKTKE---TKNEATSYQQSDKNSQTPNSDNV 607
Qy 264 LKONIILKRGYALKV-----TTYNPGKTDMLSGNGVYSKED- 299
Db 608 IKKNMIDRKTCSIKYVVLIPNNINHEIINKLILINSTYKE-----MYEKENL 657
Qy 300 ----IAKIQKANPRLALSETTIYADSRNVEDGRSTQSVL-MSALDGFN---IIRYO--V 349
Db 658 LPDTIYLIQ--SKSISSEIKKYMIISIKCVNGIYLYPILENVLDNFKNPMINSQNN 715
Qy 350 FTFK-----MNDKGEAIDKD-GNLVTDSSKLVLFGRKDDKEYTGEDKFNFVAEKDGSM 402
Db 716 STFPTVSVLYNDE---IDREIGNLJ-NTNKLJ-----NKIKIKLPEKGYI 759
Qy 403 -----FIDTKPVNLSMDKN-----YFNPSEKSNKIY 427
Db 760 INDICFRFPDKNNVLVCVDRNRKNSGNGIQONLEIDHIGHVETANYIFNLMSNFII 819
Qy 428 VRNP-----EFLRGKISDKGFNWEELRVNESVVDNVLIIYGDLDHIDN 469
Db 820 VKSSLDLFPSSHIFPLIEQVYVDFIEKNELLLKLFHFLAITAICINLCFL----- 872
Qy 470 TRDFNIKLVNKGDDIMDMGMDKYKAGFPDKVTDMDGNVYLOTGYSDLNAKAVGHYQFL 529
Db 873 ---FNINAQI---DSLFI---KNLKAQVQKDEQNNHFTFLYTSYEB-RVQNNGKHCPYF 923
Qy 530 YDNVNP-----EWNIDPKNNTSEIYAD--GKSVVFNINDKRNNGDFGEIEQHIYI 578
Db 924 LNTYSKPLIFRQSENELN-----NISHADVEGHTNLGNTSDSHRIKCEBEEIKKKKIE 976
Qy 579 NGKEYTSFNDIKOIIDKTLNLIKIVVDFARTTVKPEFILNKDGTGEVSELKPHRVTVTIQN 638
Db 977 NG-----IK-LLKKTNDNDFIYDYARKQSI--FLYN-----AIKNDDAIKNEN 1018
Qy 639 GKEMSSTIVSEEDF-----ILPVYKGELEK-----GYQF-----DGW 670
Db 1019 NYDKKKQIFVTHSYTKKQFQDYVINLRYNQLQKEKHEDYSFRREVETLQSNIRDQ 1078
Qy 671 EISGFEGKKDAG-YVINLSKDTIKVPFKLIEKKEE-----ENKPTFDYSKKDNQVNV 724
Db 1079 DVKSYFGKNNNIFSGGLNQ-----FDCKNEKKREGIIGKSSNNKSWQNDQNTDYNQNS 1131
Qy 725 HSQLNESHKEDLQREHSQKSDSTKDVATVLDKKNISSKSTNN 770
Db 1132 KQTFSDKKFQNIYTDENIQNLN-----NNFSTNQVNN 1165

RESULT 13
ID Q8I3P4_PLAF7
AC Q8I3P4_PLAF7 PRELIMINARY; PRT; 1777 AA.
DT 01-MAR-2003, integrated into UniProtKB/TrEMBL.
DT 01-MAR-2003, sequence version 1.
DT 07-FEB-2006, entry version 11.
DE Hypothetical protein PFE1095w.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2006, 22:49:27 ; Search time 22 Seconds
(without alignments)
792.558 Million cell updates/sec

Title: US-10-067-385-8

Perfect score: 4026

Sequence: 1 KLGETAESKFKVLNGKEGS.....ATVLKNNISSKSTTNNPNK 773

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 96747 seqs, 22556637 residues

Total number of hits satisfying chosen parameters: 96747

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA_New.*

- 1: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 2: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 3: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 4: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 5: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 6: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 7: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 8: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US60_NEW_PUB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	186	4.6	1349	6	US-10-471-571A-3352
2	181.5	4.5	995	6	US-10-471-571A-3354
3	177.5	4.4	991	7	US-11-165-586-62
4	177.5	4.4	1274	6	US-10-471-571A-998
5	162.5	4.0	1169	6	US-10-527-411-20
6	159	3.9	809	6	US-10-471-571A-5394
7	154	3.8	1003	6	US-10-471-571A-4826
8	153	3.8	719	6	US-10-449-902-45970
9	151.5	3.8	981	6	US-10-471-571A-666
10	150	3.7	1217	6	US-10-471-571A-4942
11	149.5	3.7	872	6	US-10-527-411-145
12	147	3.7	879	6	US-10-527-411-143
13	147	3.7	887	6	US-10-527-411-147
14	147	3.7	895	6	US-10-471-571A-922
15	147	3.7	1261	6	US-10-471-571A-1304
16	147	3.7	1315	6	US-10-527-411-141
17	147	3.7	1517	7	US-11-257-500-9
18	147	3.7	1526	7	US-11-257-500-15
19	146	3.6	876	6	US-10-527-411-128
20	146	3.6	877	6	US-10-527-411-126
21	146	3.6	877	6	US-10-527-411-130
22	146	3.6	881	6	US-10-527-411-124
23	146	3.6	902	6	US-10-527-411-132
24	146	3.6	912	6	US-10-527-411-116
25	146	3.6	914	6	US-10-527-411-120

26	146	3.6	944	6	US-10-527-411-122	Sequence 122, App
27	146	3.6	950	6	US-10-527-411-118	Sequence 118, App
28	146	3.6	1009	6	US-10-471-571A-2402	Sequence 2402, App
29	144.5	3.6	651	6	US-10-471-571A-252	Sequence 252, App
30	142.5	3.5	682	6	US-10-471-571A-992	Sequence 992, App
31	142.5	3.5	1067	6	US-10-521-401A-14	Sequence 14, Appl
32	141.5	3.5	682	6	US-10-521-401A-15	Sequence 15, Appl
33	140	3.5	449	6	US-10-471-571A-1756	Sequence 1756, App
34	138.5	3.4	995	6	US-10-471-571A-3910	Sequence 3910, App
35	138.5	3.4	1842	6	US-10-511-937-2929	Sequence 2929, App
36	136.5	3.4	2871	6	US-10-505-928-100	Sequence 100, App
37	135.5	3.4	9535	6	US-10-471-571A-4496	Sequence 4496, App
38	134.5	3.3	855	6	US-10-521-401A-11	Sequence 11, Appl
39	134.5	3.3	1066	6	US-10-521-401A-13	Sequence 13, Appl
40	133.5	3.3	512	6	US-10-471-571A-2412	Sequence 2412, App
41	133.5	3.3	727	6	US-10-471-571A-5580	Sequence 5580, App
42	133.5	3.3	774	6	US-10-471-571A-4824	Sequence 4824, App
43	133.5	3.3	858	6	US-10-527-411-22	Sequence 22, Appl
44	133.5	3.3	1328	6	US-10-504-973-32	Sequence 32, Appl
45	133.5	3.3	2478	6	US-10-471-571A-2278	Sequence 2278, App

ALIGNMENTS

RESULT 1

US-10-471-571A-3352
; Sequence 3352, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927WO
; CURRENT APPLICATION NUMBER: US/10/471,571A
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: GB-0107661.1
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 3352
; LENGTH: 1349
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (1)-(1349)
; OTHER INFORMATION: Ser-Asp rich
US-10-471-571A-3352

Query Match	4.6%	Score 186;	DB 6;	Length 1349;
Best Local Similarity	20.8%	Pred. No. 0.03;		
Matches 188;	Conservative 115;	Mismatches 328;	Indels 274;	Gaps 51;
QY	23	KDTTGVHHQENE-----ESIKKSSFTIDRNISTIRDFENKOL---KK 64		
Db	164	KSVNVQPTNEENKVKDAKTESTTLNVKSDAIKSNDETLDVNN-SNSNNENNADILPKS 222		
QY	65	LIKKKF-----REVDD-FTSETGKMEEDYKYDDKGN-IIAYDDGTDL 106		
Db	223	TAPKRLNTRMRIA AVQPSSTEAKNVNDLITSNTLT TVVD-----ADKKNKIVPAQDYLSL 277		
QY	107	EYETEKLDEIKS-----KIYGLSPSKDGHFEILLGKISNVSKNAKVYGGNNYK 154		
Db	278	KSQITVDVKVSGDYFTIKYSDTVQVYG-INPE-----DIKNI---GDIKDPNNGE 324		
QY	155	SIEIKATKYDFHSKTMTF---DLYANINDIVDGLAFA-----GDMRLFPVK-DN 198		
Db	325	T--TATAKHDVANNLITYTFTDYVDRFNSVQMGINSYIMDADTTPVSKNDVFEVNTIGN 382		
QY	199	DQKAEIKIRMPKIKETKSEY-----PYVSSYGN-----VIELGEGDLSKNK- 241		
Db	383	TTTTTTANIQYDPDYVNVNEKNSIGSAFTETVSHVGNKENPGYKYQTIYVNPNSLNAKL 442		

[illegible]


```
Db 30 EFRSVDKHOIADVTNVPDYEK-----LRNTWLDVNYGYDKYDE-----NN 72
QY 126 PSKDGHEIFLGK-ISNVSKNAKYVYGNVYKSIKATKYDFHSKWTMTFDLYANINDIVG 184
Db 73 PDMKKFDATEKATNLLKEMKTESGRKY--LWSGAETLETNSHMT-RTYRNIKIAEA 129
QY 185 LAFAGDMRFLFVKDNDQKQKAEIKIRIMPEKIKETKSEYPYVSSYG-----NVIELGEG---D 236
Db 130 MR---NPKTTLTNDENK-----KVKDA-LEWLHKNAVYKPEDKVKVKELSNFNFTKT 176
QY 237 LSKNKPDLNLTWESGKIYSDSEKQOVLKDNIL-----RKGYALKVTTYNP-----283
Db 177 TGKNTNLNWDVEIGTPKS-----LNTNLTLLNDQFSNEEKKKFTAPIKTFAPDSDK 228
QY 284 -----GKTDMLEGNVYSGKEDIKIOKANPNLRALSETTIYADSRNVEDGRSTOSVLSA 338
Db 229 ILLSVGKAEALAGGNLV---DISVKV-----LLECIIEED-----KDMKKKS 267
QY 339 LDGFNIIRYQVFFKNDKGEAIDKDGNLVTDSSKLVLFGKD-----DKSYTGEDKFN 391
Db 268 IDSFN---KVFTY-VQDSATGKERNG-----FYKDGSIYDHQDVPYTG---AVG 309
QY 392 VEAIKEDGSM-LFIDTKPNVLSMDKNYFNPSKSNKIYVN-----PEFYLRGKISDKG 443
Db 310 VVLEGISQMPMIKETP-----FNDKTQNDTTLKSWIDDGFMPLYI-KGENMDLS 359
QY 444 GPNWELRVNE-----SVVDNVLIVGDLHIDNTRDFNKLNVKDGIDMDWGMKDYKANG 496
Db 360 RGRALSRENETHSASATVWKSLLRLSDAMDSTK-AKYKKIYKSVESD-----SSYKQND 415
QY 497 FPDKVTMDMGVYLTQGYSDLNKAVGVHYQFLYDNVKNPEVNI DPKGNTSIEYADGKSVV 556
Db 416 YLNSYSDIDKMSLMT---DNSISKNGLTQQLKIYNDMDRVYH---NKDLOFAGLSMT 469
QY 557 FNINDKRNNGPGEIQOEHIYNGKEYTSFNDIKQIIDKTLNIKIVVDKDFARNTTKPEI 616
Db 470 -SKNVARYESINGENLKGWHTGAGMSYLYNSDVGHYHD-NFVWVTADMKRLSGTTLIDNEI 527
QY 617 LNKDGTGEVSELPHRVTVTIQNG-----KEMSTIVSEEDFIL-----PVYK 658
Db 528 L-KDTDKSSK-----TFVCGTKVDHDSHAGHMFENQDKTLTAKSYFINDKIVFL 580
QY 659 GELEKGYQPDGWEISGFEKKDAGYVINLSKDTFIKPVFKKIEEKEENKPTF---DV 714
Db 581 GTGIKSTDSKPNVTTIENRKANGVLYTDD-----KQTTNSDNQENNSVFLESTDT 632
QY 715 SKKKNPQVNHSQL---NESHK---KEDLQREHSQKSDSTKDVATVLDKNNISKSTT 768
Db 633 KKNIGYHFLNKPKITVYKESHTGKWKINKSKQDKTKTDEYYEVT-----QKHS 681
QY 769 NNPKN 773
Db 682 NSDNK 686
```

```
RESULT 7
US-10-471-571A-4826
; Sequence 4826, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927W0
; CURRENT APPLICATION NUMBER: US/10/471,571A
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: GB-0107661.1
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 4826
; LENGTH: 1003
; TYPE: PRN
```

```
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (1)-(1003)
; OTHER INFORMATION: chromosome segregation SMC
US-10-471-571A-4826

Query Match 3.8%; Score 154; DB 6; Length 1003;
Best Local Similarity 18.4%; Pred. No. 0.9;
Matches 157; Conservative 146; Mismatches 335; Indels 214; Gaps 39;

QY 5 IAESKPNLGNKGESGLKKOT--TGVEHHHQNESIK-----EKSSFTIDRN--ISTIR 55
Db 49 LGEQSAKSL---RGSKMEDIIFSGAEHRKAQYAEVOLRDNHSHKSLSDVENEVIVTRR 104
QY 56 DFNKDLKLI---KKKPREVDFTSETCKRMEY-----DYKYDDKGNIIAY 100
Db 105 LYSRSESEYYINNDRARLKDIADFLDGLGLGKEAYSIISQGRVDEILNKPIDRRQIIEE 164
QY 101 DDGTDLEYE-----TEKLDEIKSIYGVLSPSKDGHEILGKISNVSKNAKYVYGNVYK 154
Db 165 SAGV-LKYKRAKAEISLKLQDQEDNLRV---EDILYDLEGRVPELKEAAI--AKYK 217
QY 155 SIEIKATKYDFHSKWTMTFDLYANINDIVDGLAFAGDMRLFVKDNDQKQKAEI-KIRIMPEKI 213
Db 218 TLHQMKHSDIVVTVHDIDQVTDNDRQL-----DQRLNDLQGGQANKKADKQRLSQOI 270
QY 214 KETKSE-----YPIVSSYGNVIELGEGDLKSNKPNLNTWESGKIYSDSEKQO 261
Db 271 QQYKGRHQLDNVDSLNYQLVKA-TEAFKTYTQGLNVLEERKKNQSQSTNARYEEQENL 329
QY 262 YLLKDNILIRKGYALKVTTYNPDKTDMLEGNVYSKEDIKIOKANPNLRALSETTIYAD 321
Db 330 MELLENI-----SNEISEAQDTY-KSLSKOKELNAVIRELEEQLYVSD 372
QY 322 ---SRNVEDGRSTQSVLMS-ALDGFNIIRYQVFFKMDN-KGEAIDKGNLVTDSKLYL 376
Db 373 EAHDEKLEIKNEYTYTLMSEQSDVNDIRFLKHTTEENEAKKSRLD-----SRLYE 423
QY 377 FKODKEYTGEDKFNVEAIKEDGSMFLFDTKPNVLSMDKNYFNPSKSNKIYVNRPEFYLR 436
Db 424 VPEQKDIQGGQIKTKTKYQQTNKELSADVDEIK-NIEKDLTDTKKAQNEY--EEKLYQA 480
QY 437 GKISDKGFNWEELRVNESVVDNVLIVGDLHIDNTRDFNIK-LNVKDGIDMDWGMKDYKAN 495
Db 481 YRYTEKMTRIDSLATQEEYTYFPNGVKHILKAKNELKGIYGAVAEIID-----531
QY 496 GPPDKVTMDMGVYLTQGYSDLNKAVGVHYQFLYDNVKNPEVINIDPKGNTSIEYAD---551
Db 532 -VPSKLTQA---IET-----ALGASLQHV-----VDSEKDGQRAIQFLKERNL 571
QY 552 GKSUVFNIN-----DKRNNGPGEIQOEHIYNGKEYTSFNDIKQIIDKTL 597
Db 572 GRATFLPLNVIQSRVVATDIKSIKANGP-----ISIASEAVKVAPEYQNTIGNLL 623
QY 598 NIKIVWDXDFARNTTVKEFILNKDTGEVSELKPHRV-TVTIO-----NGKEMSS 644
Db 624 GNTIIVDHL-----KHANELARAIKYRTRIVTLEGDIVNPGSGMTGGARKSK 671
QY 645 TIVSEEDFILPVYKGELEKGYQFDGWEISGFEKKDAGYVINLSKDTFIKPVFKKIEEKK 704
Db 672 SILSKQD-ELTTMRHQLEDYLR---QTESFEQO-----FELKTKS 708
QY 705 EENKPTFDVSKKD--NPQVNHSQL-----NESHKEDLQREHSQKSDSTKDVAT 755
Db 709 DQLSLEYFEKSKQNTLKEQVHHFEMELDRLTQTQIKNDHEEFEFEKNDGYTSDKS--766
QY 756 VLDKNNISSKST 767
Db 767 ---RQTLSEKET 775
```

```
US-10-449-902-45970
; Sequence 45970, Application US/10449902
; Publication No. US20060123505A1
; GENERAL INFORMATION:
; APPLICANT: National Institute of Agrobiological Sciences.
; APPLICANT: Bio-oriented Technology Research Advancement Institution.
; APPLICANT: The Institute of Physical and Chemical Research.
; APPLICANT: Foundation for Advancement of International Science.
; TITLE OF INVENTION: FULL-LENGTH PLANT cDNA AND USES THEREOF
; FILE REFERENCE: MOA-A0205Y1-US
; CURRENT APPLICATION NUMBER: US/10/449,902
; PRIOR FILING DATE: 2003-05-29
; PRIOR FILING DATE: 2002-05-30
; PRIOR FILING DATE: 2002-203269
; PRIOR APPLICATION NUMBER: JP 2002-383870
; PRIOR FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 56791
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 45970
; LENGTH: 719
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-449-902-45970

Query Match      3.8%; Score 153; DB 6; Length 719;
Best Local Similarity 19.3%; Pred. No. 0.64; Mismatches 292; Gaps 39;
Matches 154; Conservative 104; Indels 249;

Qy 72 EVDDFTSETGRMEEDYKVDKGNIIAYDDGTLEY-----ETEKLD 114
Db 25 DLDSVTLPAG---EDFGILSDDE-DLLRIDPMELENGFANVVVDNLVPVPPKFEKLE 80
Qy 115 EIKSKIYGLSPSKDGHFEILGKISNYSKNAKVYVG-----NNYKSIEIKATKYDPSHK 168
Db 81 NVIRKISQIGAIKEG-----GLWVPVNPETKTYGYCFEYNTPOEALAREKTINGYKL 135
Qy 169 TMTEDLVANINDIVDGLAFAGDMFLVKDNDQ-KKAEIKRMPEK-----IKETKSEVPY 222
Db 136 DKSHIFAVNMFED-----DFQYMKVPDEMMPAEIKPYTPGENIKQMLADEKARDQF 186
Qy 223 VSSYGNVIELGEGDLSKNKPNLTKMESGKIYSDSEKQ-----OYLL----- 264
Db 187 VIRAGTFEVYNDARQKMPQLVFKQ---YWTDSFIQWSPLGYLATVHRQGSQVWGGD 243
Qy 265 -KONILIRKGYA-LKVTYTPNGKTDMLEGNVYSKEDIKQKPNLALSETTIYADS 322
Db 244 DKFERLARFAHAQVKLIDFSPEKYL---TYSSH-----EPSNP----- 280
Qy 323 RNVEDGRSTQSVLMSALDGFNIIRYQVTFK-MND-KGEA--IDKGNLVTDSKLVLF- 377
Db 281 -----RDTHRVLNIFD-----VRTGKVMRDFKGSADFPASGNSISVSGVSWFIFR 326
Qy 378 ---GKDDKEYTGEDKFNVEAIKEDGSLFIDTK----- 407
Db 327 WGGGKDKYFARLCK-NVISVYETETSLDDKSKLVENVVDFCWSPTDPIIALFVPELG 385
Qy 408 ----PVNLSM-----DKNYFNPSKSNKIYVRNPEF-----YLRKISDKGGFN 446
Db 386 GGNQPARVSLVQIPGKBELRQNLFSVSDCKWYQNSGEYLAQVDRYTKTKSTYTGFE 445
Qy 447 WELRVNESVVDNYLIYGDLDHNDTDFNKLNVK-----DG---DINDWGMKDY 492
Db 446 L-FRIKERDPIEV-----LELDNKNKIIAFAPWEPKGRFRAVHGDGPKPDISYSMTKS 500
Qy 493 KANGFPKQVTDMDGNVYLQTYGSDLNAKAVGVHYQFLYDNVKEPVNIDPKGNTSIEYADG 552
Db 501 NNISRVSKLITLKGK----- 526
Qy 553 KSVFVNINDKRNNFGDEIQEIQHIYINGKEVTSFNIDKIIQIDKTLNIKIVVKDPAFRTTV 612
Db 527 RF1VF-----AGLKGFGNGL-----EFYNNVDELETWA----- 553
Qy 613 KEFILNKDTSVSELKPHRVTVTIQ---NGKEMSSTIVSEEDFILPVYKGELEKGYQFDG 669

US-10-471-571A-666
; Sequence 666, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927WO
; CURRENT APPLICATION NUMBER: US/10/471,571A
; CURRENT FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: GB-0107661.1
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: Seqwin99, version 1.03
; SEQ ID NO 666
; LENGTH: 981
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (1)..(981)
; OTHER INFORMATION: conserved hypothetical
US-10-471-571A-666

Query Match      3.8%; Score 151.5; DB 6; Length 981;
Best Local Similarity 18.6%; Pred. No. 1.2; Mismatches 274; Gaps 39;
Matches 150; Conservative 136; Indels 245;

Qy 18 EGSLLKDDTTGVEHHHQENESIKEKSFITDRNIST-RDPFNKDLKKLKKKFKFV--DD 75
Db 198 ESQIRREEAKLETVRLVDD--RDKSS---RRLNKLKHLNQ-LSNWHEEKQKEVALHD 250
Qy 76 FTSETGRMEEY----DYKYDDKGNIIAYDDG-----TDLEYETEKLEIKSKIYGV 123
Db 251 HSQE-WKSLEQQLNIEPIITPPEKG-VDRYEKARAHKQSLERDIGLRNERLAQLKEEA-TQ 307
Qy 124 LSPSKDGH---FEILGKISNVSKNAKVYVGNVYSIEIKATKYDFHSKTMFTD-LYANI- 178
Db 308 LEPVQSDIDAFISLNOQENEIKV-----KEFELTAIEKDIANKQRDKDELOSIG 358
Qy 179 -----NDIVDGLAFAGDMFLVKDNDQKAEIKIRMEPEKIKETKSEYPYVSSYGNVIELG 233
Db 359 WSETHHDV-----DSSEA-MKSYVSEQIKNKQEQAAIYKQL----- 393
Qy 234 EGDLSKNKPNLTKMESGKIYS--DSEKQVLLKDNILIRKGYALKVTKYTPNGKTDMLG 291
Db 394 ERSLEEN-----KIEDNAVHSELDVSEKIVPPEETFEKKKEYSQOYIELN-----EK 440
Qy 292 NGVYS--KEDIAKIQKANPNLALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQV 349
Db 441 ENLYSKLKEFEFEIQEQKQKRLRTYF-----ILLTVGIGL---TA 481
Qy 350 FTFAMNDKGBAIDKDGNLVTDSSKLVLFG-----KDDKEYTGEDKFN 391
Db 482 FSPISNN-----MLFGIIFAVLTLVFVIGIIMSKEVDYSEAITDE 523
Qy 392 VEAIKEDGSLFIDTKPVNLSMDKNY-----FNPSSKNKIYVRNPEFVLRKISDKGGFN 446
Db 524 IBEIKAQAIL-LDENYDLDFDLDEQYRIRDHWOQALKNKQIDLEEKQYIEGLNDAKGRH 582
Qy 447 WELRVNESVVDNYLIYGDLDHNDTDFNKLNVKGOIMDWGMKDYKANGFPDPKVTMDMG 506
```

Db 583 DEL---QSTVE-----NPKD----- 594
QY 507 NVYLOTGYDLNAAVGVHYQFLYDNVKNPEVNIIDPKGNTSIEYADGKSVVFNINDKRNNG 566
Db 595 ELYLSKSN-----DLIVDSISTMANIKALDQHSIDLNQORQOLVOELDTFFYNH 644
QY 567 FGEEIOEQHYINGKEYTSFNIDKIQIIDKTLNIIKIVVDFARNTTVKEFILNKDTGEVSE 626
Db 645 ABAAVTKSQPVYFN--KLSLFHDVQOVLKS-----ABDTNEKWRINAENTKLVN 691
QY 627 LKPHRVTVTIQNGKEMSS-----TIVSEEDFPLPVYKGELEKYQPDGWEISGFEG----- 677
Db 692 ELNHLNAQLEENKKEITALDFINVGTEEDFY-----QHHDYQYTTSNLSRNFNDLTKY 745
QY 678 KKDAGYVNLSDTKFLPKVFKIEE-----KKEENKPTFDVSKKD--NPQVNHSQLN 729
Db 746 LENQVSYELSSLSSEKTTAQLLEEDHLLATQVDEYNEQVLEMAQVSDLSAQINMETD 805
QY 730 ESHRKEDLQREHS---QKSDSTKD 751
Db 806 TT--LANLRHEYSLKNQNLNDIAKD 828

RESULT 10

US-10-471-571A-4942
; Sequence 4942, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927W0
; CURRENT APPLICATION NUMBER: US/10/471,571A
; PRIOR FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: GB-0107661.1
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: Seqmin99, version 1.03
; SEQ ID NO 4942
; LENGTH: 1217
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1217)
; OTHER INFORMATION: hypothetical protein
US-10-471-571A-4942

Query Match 3.7%; Score 150; DB 6; Length 1217;
Best Local Similarity 20.4%; Pred. No. 1.9;
Matches 180; Conservative 135; Mismatches 320; Indels 248; Gaps 46;

QY 39 IKEKSSFTIDRNISTIRD--FENKDLKKLKKK-----PREVDDFTSETGR----- 83
Db 289 IKEANEMVD-----ALEDAKLYKKYSLIDKVSDFSRADDLKADMQQLAPRVKYLA 344
QY 84 -----MBEYDYKDDKGNIIAYDDGTDLLEYETEKLDSEIKSIYGLVSPSKDG-----HF 132
Db 345 RIVKDMVFENFKKSK-NILDFSD-----YEHFAL-QILTNEGD--SPSEIAESYRQHF 395
QY 133 E-----ILGKISNVSKNAKYVYNNYKSIETKATKYDHF--SKTWTFDLYANI 178
Db 396 QEBILVDEYQDTRNVQEKILSCITGDEHNGNLFMVGDVKQSIYKPRQADPSLFIEKYQRF 455
QY 179 NDIVDGLAFAGDMRLFVKNDQKAEI-----KIRPEKIKETKSEYPPVSSYGVNI 230
Db 456 TIDGGGTG-----RRIDLQNFRRSRKEVLSTTNIFKHMWDEQGVGEVYDEAAQLYYGAPY 511
QY 231 ELGEDSLSKNPDNL--TKMESKIIYSD---SEKQOYLL-----KDNILRKGYALKVTTYN 282
Db 512 D-----ESDHPVNLKVLVEADQEHSDLTGSEQEAHFIVEQVKDILEHQKVDYDMKTGYSR 565
QY 283 PGKTDMLGNGVYSKEDIATQKANPNLRALSETTIYAD---SRNVEDGRSTQSVL----- 335

Db 566 SA-----TKDIVILERSFGQARNLQOAFKNEDIPPHVNRSREGYFQTEVRVLV 614
QY 336 --MSALDG-----FNIRYQVFTFKMNDKCEADKDGNLVTDSSKLVLFKDDKEYTGG 386
Db 615 SPFLRAIDNPLQDIYLVGLMRSYVIYQFKDELAQI-----RILSPNDDDYFYQS 661
QY 387 -BDKFNVEAIREKDGSMFLFDTKPVNLSMDKNYFNFPSKSNKIYVRNPEF-----YLRG 437
Db 662 IWVYINDEA---ADAILVDLKMFLSDIQSQYQYKSDHPVYQLDKFNDHVVIOYFSG 717
QY 438 KISDKGG-----FNWELRV-----NESVV---DN--- 458
Db 718 LIGGRGRANLYGLENFKAIEFENSFRGLYQFIRFIDELIERGKDFGEENVVGPNDNVVR 777
QY 459 -----YLIYGDHLID-NTRDFN--IKLVNKGDDIMDMGMDKYKANGPDKVT 502
Db 778 MMTIHSSKGLBPPFFYIYSGLSKDFNKRDLKQPVILNQOQFGLGMDYFDVD-KEMAF----- 832
QY 503 DMDGNVYLQTYGSDLNAAKAVGVHY-----QFLYDNVKNPEVNIIDPKGNTSIEYADGK 553
Db 833 SLASVAYRAVAEKELVSEEMRLVYVALTRAKEQLYLGVRKNDKSLLELEQLSI---SGE 889
QY 554 SVVFNINDKRNNGFD--GEIOEQHYIINGKEYTSF-NDIKQIIDKT---LNKIVVVKDPA 607
Db 890 HIAVNERLTSPNPHLIYSILSKHQSASTPDDLKEKDIAQIEDSSRPNNVISIV---YF 946
QY 608 RNTTVKEFILANKDTGEVSELKPHRVTVTIQNGKE--MSSTIVSEEDFILPVYK----- 658
Db 947 EDVSTETILDNDYKSVNQL-----TWQNGEDVKAQIKHOLDYRYPVNDTKPKSKQ 1000
QY 659 --GELEKGYQFDGWEISGFEGKKDAGYVYINLSKDTFIKPVFKKIEEKEENKPTFDVS- 715
Db 1001 SVSELKQRYE---TRESGTSYRVQYRIGFS--TYERPKEFSEOGKKANEIGTLMHTV 1055
QY 716 -----KKKDNPPQVNHSQLNESHRKEDLQREHSQKSDSTKDV 752
Db 1056 MQLHPFKKERISEV-----ELHQYIDGLIDKHIIEADAKKDI 1092

RESULT 11

US-10-527-411-145
; Sequence 145, Application US/10527411
; Publication No. US20060110410A1
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford
; APPLICANT: Foster, Keith Alan
; APPLICANT: Chaddock, John
; APPLICANT: Marks, Philip
; APPLICANT: Sutton, J. Mark
; APPLICANT: Stancombe, Patrick
; APPLICANT: Wayne, Jonathan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130005
; CURRENT APPLICATION NUMBER: US/10/527,411
; CURRENT FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: PCT/GB2003/003824
; PRIOR FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: US 10/241,596
; PRIOR FILING DATE: 2002-09-12
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 145
; LENGTH: 872
; TYPE: PRT
; ORGANISM: Clostridium tetani
US-10-527-411-145

Query Match 3.7%; Score 149.5; DB 6; Length 872;
Best Local Similarity 20.4%; Pred. No. 1.3;
Matches 131; Conservative 107; Mismatches 230; Indels 175; Gaps 36;
QY 79 ETGRMEEYD-----YKYDDKGNIIAYDDGTD--LEYETEKLDSEIKSIYGLVSPS 127

[illegible]

RESULT 12

US-10-527-411-143
; Sequence 143, Application US/10527411

```

, Publication No. US20060110410A1
,
, GENERAL INFORMATION:
,
, APPLICANT: Shone, Clifford
,
, APPLICANT: Foster, Keith Alan
,
, APPLICANT: Chaddock, John
,
, APPLICANT: Marks, Philip
,
, APPLICANT: Sutton, J. Mark
,
, APPLICANT: Stancombe, Patrick
,
, APPLICANT: Wayne, Jonathan
,
, TITLE OF INVENTION: Reconstituted Toxin Fragments
,
, FILE REFERENCE: 1581_0130005
,
, CURRENT APPLICATION NUMBER: US/10/527,411
,
, CURRENT FILING DATE: 2005-03-11
,
, PRIOR APPLICATION NUMBER: PCT/GS2003/003824
,
, PRIOR FILING DATE: 2003-09-12
,
, PRIOR APPLICATION NUMBER: US 10/241,596
,
, PRIOR FILING DATE: 2002-09-12
,
, NUMBER OF SEQ ID NOS: 175
,
, SOFTWARE: PatentIn version 3.2
,
, SEQ ID NO 143

```

Query Match 3.7%; Score 147; DB 6; Length 879;

[illegible]

RESULT 13

```

RES001 13
US/10-527-411-147
; Sequence 147, Application US/10527411
; Publication No. US20060110410A1
; GENERAL INFORMATION:
; APPLICANT: Shone, Clifford
; APPLICANT: Foster, Keith Alan
; APPLICANT: Chaddock, John
; APPLICANT: Marks, Philip
; APPLICANT: Sutton, J. Mark
; APPLICANT: Stancombe, Patrick
; APPLICANT: Wayne, Jonathan
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581_0130005
; CURRENT APPLICATION NUMBER: US/10/527,411
; CURRENT FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: PCT/GB2003/003824
; PRIOR FILING DATE: 2003-09-12
; PRIOR APPLICATION NUMBER: US 10/241,596
; PRIOR FILING DATE: 2002-09-12
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 147
; LENGTH: 887
; TYPE: PRT

```



```
; ORGANISM: Clostridium tetani
US-10-527-411-147

Query Match          3.7%; Score 147; DB 6; Length 887;
Best Local Similarity 20.0%; Pred. No. 1.8; Mismatches 182; Gaps 36;
Matches 130; Conservative 110;

QY 79 EFGKRMEEYD-----YKDDKGNIIAYDDGTD--LEYETEKLEIKSIYGVLSPS 127
Db 51 EFGTKPEDFNPSLLLEGASEYDIP-NYLRTSDKDRFLQTVKVLFNRIKNNVAGE--- 105

QY 128 KUGHPEILKINSVSNKAVYGNKYSEIKATKYDFHSHKWTWFDLY-ANINDIVGLA 186
Db 106 -----ALLDKII-----NAIPYLGNSVLLD---KPDNTNSVSNFLLQDPDSGATTKSA 152

QY 187 FAGDMELFVKDNDQKAEIK--IRMEPIKETKSEYP-----YVSSVG 227
Db 153 MTLNLIIFGPGVNLKNVGRGIVLRV-----DNKNYFCRGGFSGIMQWAFCEYVPTFD 207

QY 228 NVIELGEGDLSKNKPNLTKMBESGIYSDSEKQQVLLKDNII-ILRKGVALKVTYTN--PG 284
Db 208 NVIE-----NITSLTIGSKSYFQDPALLMLHLELIHVLHCLYGMQVSSHEIIPS 255

QY 285 KTDMLE-----GNGVYSK-----EDIAKIOKAN-P 308
Db 256 KQEIYMQHTYPIASBELFTFGGQDANLISIDIKNDLYEKTLDNYKAIANKLSQVTSNDP 315

QY 309 NLRALSETTIYADSRNVEDGRSTQSVL-----MSALDGFNII-----RYQVET---- 351
Db 316 NIDIDSYKQIYQOYQFQKDSNGQYIVNEDKQFQILYNSIMYGFTEILGKFNKIKTRLISY 375

QY 352 FRMDKGEAIDK--DGNLVTDSKVLFGKDDK-EYTGED-RFNVEAIKE-DGSMFLDT 406
Db 376 FSNMHPVKIPNLLDDTIYNDTEGNIIESKDLKSEYKQNMVNTNAPFNVDSGLV--S 433

QY 407 KPNVL-----SMDKNVFNPSKNK-----IYVRNPEYLRGKISDKGGFNWELRV 451
Db 434 KLIGLCKKIIPPTNIRENLNRYNTASLTDLGELCIIKKNEDLTF---IAEKNSFS----- 485

QY 452 NESVDNLYLYDGLHDNTRDFNKLNVK-DGDMWGMKDYKANGFPDKVTDMDGNVYL 510
Db 486 BEFPQDEIVSY-----NTKNKPLNFNYSKDKIIVDYNLQS-KITLPNDRTTPVTKGIPY 538

QY 511 QYGYSDLNKAVGVH-----YQFLYDNVKEPV-----NIDPK-GNTSIEYADGK 553
Db 539 APEYKSNAASTTEIHNIDNTIYQIYAKSFTTLQRTWNSVDALINSTKI-YSYFP 597

QY 554 SVFVNINDKRNNGFDGEIQEHIYINGKEYTSFNDIKQIIDKTLNIIKIVV 603
Db 598 SVI-----SKVNOGAQGLFLQWVRDIIIDFTNESQKTTIDKISDVSTIV 643
```

```
RESULT 14
US-10-471-571A-922
; Sequence 922, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927WO
; CURRENT APPLICATION NUMBER: US/10/471,571A
; PRIOR FILING DATE: 2003-09-12
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 922
; LENGTH: 895
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1) .. (895)
```

```
; OTHER INFORMATION: hypothetical protein
US-10-471-571A-922

Query Match          3.7%; Score 147; DB 6; Length 895;
Best Local Similarity 18.5%; Pred. No. 1.8; Mismatches 237; Gaps 28;
Matches 129; Conservative 87;

QY 77 TSETGKRMEEYDYKYDDKGNIIAYDDGTDLEVEYETEKLDEIKSIYGVLSPSKDGHFELIG 136
Db 392 TASTWKKEFEVGEKGLPVLSVSDSKDYAY-----TRF 426

QY 137 KISNVSKNAKY-----YGNNYKSIIEIKATKYDFHSHKWTWFDLYANINIDVGLAFAGDMR 192
Db 427 PVSNGTREVKIVSSIEYGENIHE-----DYDYTLWFAQPIITNPNDDYVDEETY-NLQK 479

QY 193 LFKVNDNDKKAIEIKRMPEKIKETKSEYYPYSSYGNVIELGEGDLSKNKPNLTKMBESGK 252
Db 480 LLAPVHKAKTLERQVYELKLOEKLPE-KYKAEYKCKLDQTRVELADQVKSATVEFEN-- 536

QY 253 IYSDSEKQQVLLKDNIIILRKGVALKVTYTNPGKTDMLGNGVYSKEDIKQKAPNLR 312
Db 537 -----VPTNDQLTDLQEAHFV----- 553

QY 313 LSETTIYADSRNVEDGRSTQSVLSALDGFNIIYQVFTFKNDKGEAIDKGNLVTDS 372
Db 554 -----VPESEENSE-----SVMDGF--VEHPFVTATLNGQKYVVMK----- 587

QY 373 KLVLFGRKDDKYEYTGEDKFNVEAIKEDGSMFLDTKPNLSMDKNVFNPSKSNKIYVRNPE 432
Db 588 -----TKDDSYMKDLIIVEGRKVTTVSKDPKN-----NSR 616

QY 433 FVLRGKISDKGGFNWELRVNSVVDNLYLYDGLHDNTRDFNKLNVKDGDMWGMKD- 491
Db 617 TLIPIYPIDKAVYNAIVKV---VVANIGYEGQYHV---RIINQDINTKDDTSDQNTSEP 670

QY 492 -YKANGFPDKVTDMD--GNVYLOQYGYSDLNKAVGVHGYQFLYDNVKNPEVNDIPKGNISIE 548
Db 671 LNVQTQGEQGVADTDVAENSSSTATNPKDASDKA-----DVIEPESDVVKDADNNID 721

QY 549 YADGKSVFVNIND-KRNNGPDGEIQEHIYINGKEYTSFNDIKQIIDKTLNIIKIVVKDFA 607
Db 722 -KDVOHQDVHLSUDMSDNNHFD-----KY-----DLKEMDTQ-----IAKOTD 757

QY 608 RNTTVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKGELEKGYQF 667
Db 758 RN-----VQKDA-----DNSVGSSNVDTDKD----- 779

QY 668 DGWEISGFEKGDAGYVINLSKDTFIKPVFKKIEBEKEEENKPT-FDVSKKKNPQVNH 726
Db 780 -----SNKNKDKVIQ--LNHIADKNNHTGKAAKLDVVKQNYN---NTD 817

QY 727 QLNESHRKEDLQREHS--QKSDSTKQVATVLDKNNIS 763
Db 818 KVTDKKTTEHLPSDIHKTVTKTKKAGTPSKENKLS 856
```

```
RESULT 15
US-10-471-571A-1304
; Sequence 1304, Application US/10471571A
; Publication No. US20060115490A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026927WO
; CURRENT APPLICATION NUMBER: US/10/471,571A
; PRIOR FILING DATE: 2003-09-12
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 5642
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 1304
; LENGTH: 1261
; TYPE: PRT
```


GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2006, 22:32:32 ; Search time 21 Seconds
(without alignments)
3541.693 Million cell updates/sec

Title: US-10-067-385-8
Perfect score: 4026
Sequence: 1 KLGEIAESKFNGLNGKEGS.....ATVLDKNNISSKTTNNPNK 773

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4026	100.0	2140	2 F95074	serine proteinase,
2	3709.5	92.1	2144	2 A97942	metalloproteinase
3	253.5	6.3	1127	2 T28317	ORF MSV156 hypothe
4	250	6.2	1650	2 T18444	hypothetical prote
5	235.5	5.8	3724	2 T18427	hypothetical prote
6	231.5	5.8	2269	2 T28677	rhostry protein -
7	230	5.7	2401	2 T28676	rhostry protein -
8	228	5.7	4550	2 T18440	hypothetical prote
9	221.5	5.5	1711	2 T18429	hypothetical prote
10	219	5.4	1640	2 A24594	probable major sur
11	218	5.4	1631	1 SAZQK1	major merozoite su
12	218	5.4	1639	2 S05603	lipoprotein (impor
13	217.5	5.4	1546	2 G90603	reticulocyte-bind
14	215.5	5.4	1252	2 B42771	DNA-directed RNA p
15	214	5.3	2339	2 A45597	membrane nuclease,
16	212.5	5.3	1125	2 E90598	hypothetical prote
17	212.5	5.3	2166	2 G70163	serine/threonine-s
18	212.5	5.3	2485	1 H71621	protein with 5'-3'
19	212	5.3	1188	2 A71621	microbial collagen
20	211.5	5.3	1104	1 A36866	RESA-H3 antigen PP
21	209	5.2	1558	2 B71603	hypothetical prote
22	207.5	5.2	4688	2 E89824	chromosome segrega
23	207	5.1	1141	2 B97116	hypothetical prote
24	207	5.1	1191	2 C71622	hypothetical prote
25	207	5.1	1979	2 A71717	probable cell surf
26	206.5	5.1	1622	2 AE1717	mature-parasite-in
27	206	5.1	1526	2 A45605	RAD2 endonuclease
28	204.5	5.1	1516	2 E71619	hypothetical prote
29	204.5	5.1	2500	2 G71609	

30	204	5.1	1245	2 D71613	GAF domain protein
31	203.5	5.1	3394	2 T18501	hypothetical prote
32	203	5.0	1166	2 T28680	fibrinogen-binding
33	203	5.0	4981	2 T18489	hypothetical prote
34	202.5	5.0	763	2 G97026	superfamily I DNA
35	202.5	5.0	1939	2 T18372	repeat organellar
36	201	5.0	1712	2 C71618	ATP-dept. acyl-CoA
37	199	4.9	888	2 E71608	hypothetical prote
38	199	4.9	980	2 E71606	hypothetical prote
39	197	4.9	1386	2 ACL533	surface protein (L
40	196.5	4.9	2523	2 T18477	hypothetical prote
41	196	4.9	1714	2 E71609	Ser/Thr protein ki
42	195.5	4.9	2380	2 E71604	hypothetical prote
43	195	4.8	1008	2 T18508	hypothetical prote
44	195	4.8	1302	1 JC6009	surface-located me
45	195	4.8	1365	2 T30822	impl protein - Myc

ALIGNMENTS

RESULT 1

F95074
serine proteinase, subtilase family [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C:Accession: F95074
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
non, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: F95074
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2140 <KUR>
A:Cross-references: UNIPROT:Q97RY6; UNIPARC:UPI000005150F; GB:AE005672; PIDN:AAK74791.1,
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0641

Query Match	100.0.0%;	Score	4026;	DB	2;	Length	2140;									
Best Local Similarity	100.0.0%;	Pred. No.	9.5e-166;													
Matches	773;	Conservative	0;	Mismatches	0;	Indels	0;									
Gaps	0;															
Qy	1	KLGEIAESKFNGLNGKEGSLKDDTTGV	VEHHHQENEE	SIKEKSSFTIDRN	ISTIRDPENK	60										
Db	1334	KLGEIAESKFNGLNGKEGSLKDDTTGV	VEHHHQENEE	SIKEKSSFTIDRN	ISTIRDPENK	1393										
Qy	61	DLKLLIKKKFREVDFTSETG	KRMEEYDYKYDDKGNIIAY	DDGTDLEYET	EKLEIKSKI	120										
Db	1394	DLKLLIKKKFREVDFTSETG	KRMEEYDYKYDDKGNIIAY	DDGTDLEYET	EKLEIKSKI	1453										
Qy	121	YGVLSPSKDGHFEILGKISNV	SKNAKVYNNYKSI	EIKATKYDP	SHSTMTF	DLYANIND	180									
Db	1454	YGVLSPSKDGHFEILGKISNV	SKNAKVYNNYKSI	EIKATKYDP	SHSTMTF	DLYANIND	1513									
Qy	181	IVDGLAFAGDMRLFVK	ONDQKAEIKIR	MPKIKETKSEY	PVSVSYGN	VI	ELGEGDLSKN	240								
Db	1514	IVDGLAFAGDMRLFVK	ONDQKAEIKIR	MPKIKETKSEY	PVSVSYGN	VI	ELGEGDLSKN	1573								
Qy	241	KPDNLTKMESGKIYSDSE	KQOYLLKNDNII	LRGYALKVTTYN	PGKTD	ML	ENG	GVYSKEDI	300							
Db	1574	KPDNLTKMESGKIYSDSE	KQOYLLKNDNII	LRGYALKVTTYN	PGKTD	ML	ENG	GVYSKEDI	1633							
Qy	301	AKIQKANPNLRAL	SETTIYADSR	NVEDGR	STQSV	LM	SALD	GFNII	RYQV	FT	FKMNDKGEA	360				
Db	1634	AKIQKANPNLRAL	SETTIYADSR	NVEDGR	STQSV	LM	SALD	GFNII	RYQV	FT	FKMNDKGEA	1693				
Qy	361	IDKGNLV	TDSSKLVLF	FGKDKKEYT	GEDKFN	VEAI	KED	GS	ML	FDIT	KPVN	LS	MDK	QY	PNP	420

Db 1694 IDKGNLVTDSKLVLPKGDKEYTGEDKFNVEAIKEDGSMFLPDTKPVNLSMDKNYFNP 1753
Qy 421 SKSNKIYVRNPEFYLRGKISDKGFWNWLAVNESVDNYLYIGDLHIDNTRDFNKLNVK 480
Db 1754 SKSNKIYVRNPEFYLRGKISDKGFWNWLAVNESVDNYLYIGDLHIDNTRDFNKLNVK 1813
Qy 481 DGDIMDMGMDYKANGPDPKVTDMGNNVLOTGYSDLNAKAVGVHYQFLYDNVKNPEVNI 540
Db 1814 DGDIMDMGMDYKANGPDPKVTDMGNNVLOTGYSDLNAKAVGVHYQFLYDNVKNPEVNI 1873
Qy 541 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEHIYINGKEYTSFNDIKQIIDKTNLINIK 600
Db 1874 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEHIYINGKEYTSFNDIKQIIDKTNLINIK 1933
Qy 601 IVVKDFARNTTVKGFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGE 660
Db 1934 IVVKDFARNTTVKGFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGE 1993
Qy 661 LEKGYPDGEWISGFEKGKADAGYVNIISKDTFIKPVFKIIEEKEEENKPTFDVSKKDN 720
Db 1994 LEKGYPDGEWISGFEKGKADAGYVNIISKDTFIKPVFKIIEEKEEENKPTFDVSKKDN 2053
Qy 721 PQVNSQLNESHKEDLQREHESQKSDSTKDVATVLDKNNISSKSTNNPNK 773
Db 2054 PQVNSQLNESHKEDLQREHESQKSDSTKDVATVLDKNNISSKSTNNPNK 2106

RESULT 2
A:97942
C:metalloproteinase (EC 3.4.21.-) A [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: A97942
R: Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, B.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A: Authors: Yang, X.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A: Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A: Reference number: A97872; MUID: 21429245; PMID: 11544234
A: Accession: A97942
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-2144 <KUR>
A: Cross-references: UNIPROT: Q8DQP7; UNIPARC: UPI00000E3490; GB: AE007317; PIDN: AAK99365.1;
A: Gene: prtA
C: Genetics:
C: Keywords: hydrolase; serine proteinase

Query Match 92.1%; Score 3709.5; DB 2; Length 2144;
Best Local Similarity 92.4%; Pred. No. 4.1e-152;
Matches 714; Conservative 26; Mismatches 32; Indels 1; Gaps 1;
Qy 1 KLGIASEKPNLGNKGLSKDVTGVEHHQNEESIEKSSFTIDRNISTIRDPENK 60
Db 1339 KLGIASEKPNLGNKGLSKDVTGVEHHQNEESIEKSSFTIDRNISTIRDPENK 1397
Qy 61 DLKKLILKKKPREVDFTSETGKRMEEDYKYDDKGNIIAYDDGTDLEYETEKLDEIKSKI 120
Db 1398 DLKKLILKKKPREVDFTSETGKRMEEDYKYDDKGNIIAYDDGTDLEYETEKLDEIKSKI 1457
Qy 121 YGVLSPSKDGHEILGKISNVSKAKVYGNYSKIEIKATKYDFHSKTMTFDIYANIND 180
Db 1458 YGVLSPSKDGHEILGKISNVSKAKVYGNYSKIEIKATKYDFHSKTMTFDIYANIND 1517
Qy 181 IVDGLAFAGDMRLVFVNDQKAEIKRMPEKIKETKSEYYPVSSYGNVIELGEGDLSKN 240
Db 1518 IVDGLAFAGDMRLVFVNDQKAEIKRMPEKIKETKSEYYPVSSYGNVIELGEGDLSKN 1577
Qy 241 KPDNLTKMESGKIYSDSEKQYLLKDNIIILRGYALKVTVTPGKTMLEGNVYSKEDI 300
Db 1578 KPDNLTKMESGKIYSDSEKQYLLKDNIIILRGYALKVTVTPGKTMLEGNVYSKEDI 1637

Qy 301 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFKNDKGEA 360
Db 1638 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFKNDKGEA 1697
Qy 361 IDKGNLVTDSKLVLPKGDKEYTGEDKFNVEAIKEDGSMFLPDTKPVNLSMDKNYFNP 420
Db 1698 IDKGNLVTDSKLVLPKGDKEYTGEDKFNVEAIKEDGSMFLPDTKPVNLSMDKNYFNP 1757
Qy 421 SKSNKIYVRNPEFYLRGKISDKGFWNWLAVNESVDNYLYIGDLHIDNTRDFNKLNVK 480
Db 1758 SKSNKIYVRNPEFYLRGKISDKGFWNWLAVNESVDNYLYIGDLHIDNTRDFNKLNVK 1817
Qy 481 DGDIMDMGMDYKANGPDPKVTDMGNNVLOTGYSDLNAKAVGVHYQFLYDNVKNPEVNI 540
Db 1818 DGDIMDMGMDYKANGPDPKVTDMGNNVLOTGYSDLNAKAVGVHYQFLYDNVKNPEVNI 1877
Qy 541 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEHIYINGKEYTSFNDIKQIIDKTNLINIK 600
Db 1878 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEHIYINGKEYTSFNDIKQIIDKTNLINIK 1937
Qy 601 IVVKDFARNTTVKGFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGE 660
Db 1938 IVVKDFARNTTVKGFILNKDTGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGE 1997
Qy 661 LEKGYPDGEWISGFEKGKADAGYVNIISKDTFIKPVFKIIEEKEEENKPTFDVSKKDN 720
Db 1998 LEKGYPDGEWISGFEKGKADAGYVNIISKDTFIKPVFKIIEEKEEENKPTFDVSKKDN 2057
Qy 721 PQVNSQLNESHKEDLQREHESQKSDSTKDVATVLDKNNISSKSTNNPNK 773
Db 2058 PQVNSQLNESHKEDLQREHESQKSDSTKDVATVLDKNNISSKSTNNPNK 2110

RESULT 3
T28317
ORF MSV156 hypothetical protein - Melanoplus sanguinipes entomopoxvirus
C: Species: Melanoplus sanguinipes entomopoxvirus
C: Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C: Accession: T28317
R: Afonso, C.L.; Tullman, E.R.; Lu, Z.; Oma, E.; Kutish, G.F.; Rock, D.L.
J. Virol. 73, 533-552, 1999
A: Title: The genome of Melanoplus sanguinipes entomopoxvirus.
A: Reference number: Z20484; MUID: 99102612; PMID: 9847359
A: Accession: T28317
A: Status: preliminary; translated from GB/EMBL/DDBJ
A: Molecule type: DNA
A: Residues: 1-1127 <AFO>
A: Cross-references: UNIPROT: Q9YVT6; UNIPARC: UPI00000F6900; EMBL: AF063866; NID: g4049647;
C: Genetics:
A: Note: MSV156

Query Match 6.3%; Score 253.5; DB 2; Length 1127;
Best Local Similarity 21.5%; Pred. No. 0.0013;
Matches 186; Conservative 155; Mismatches 338; Indels 185; Gaps 44;
Qy 7 ESFKNLGNKGLSKDVTGVEHHQNEESIEKSSFTIDRNISTIRDPENKDLK 63
Db 196 EIEFFQNDVQKIEINKQDELNKLDESKEFKKQBELNKTIDKQBELIKKLNDKEIN 255
Qy 64 KLKKKPREVDFTSETGKRMEEDYKYDDKGNIIAYDDGTDLEYETEKLDEIKSKI 111
Db 256 FNIDEKQKLLDQNSKINTLNENIKGVNLYTETKNSLNQLNEILNKDSTIKSLDEKQK 315
Qy 112 KLDEIKSKIYGVLSPSKDGHEILGKISNVSKAKVYGNYSKIEIKATKYDFHSKTMTFDIYANIND 161
Db 316 LLDDELKNNITSLVKNKSN---TKITNQQLLESSLTDFFNNANI---NINELSKKIK 367
Qy 162 KYDFHSKTMTFDIYANINDIYDGLAFAGDMRLVFVNDQKAEIKRMPEKIKETKSEYYPVSSYGNVIELGEGDLSKN 216
Db 368 LFQNDIQKANDITEQNNKITD---FPNNSTRIFKELJDTYEKKIDDIKNNNLQKLEESYK 425
Qy 217 ----KSEY----PYVSSYGNVIELGEGDLSKNKPDNLTKMESGKIYSDSEKQYLLKDNII 269

Db 426 KIDEQTEYYKINKINKEYNIDIELKNNLQKLBEEKNKIDQTEYYKINKINKEY--NDIE 483
QY 270 LRKGVALKVTTNPG-----KIDMLGNGVYKEDIAKIQKAMPNLRALSETTIYAD 321
Db 484 LKNNLQKLEENKNNKNDKLTGLKNDIESNTELFNKLNISDFDKSREIAKLN--TEYEQ 541
QY 322 SRN--VEDGRSQSVL-----MSALDGNRIYQVFTFKMDKGEADKQGNLVDSK 373
Db 542 LRKDLLENKTNELMKLSNKLSSLE-----QLYDSKKNIL-DGIDKIYNSLKEKN- 592
QY 374 LVLFGKDKKEYGDEKFNVEATKED--GSMFLFIDTKPVNLSMDKNYFN--PSKSNKIY 427
Db 593 ----DKIDYFNIEKFDIYNWENKFIKNLDSIINKIINNQQKEYINKSDSKNELS 648
QY 428 VRNPEFY-----LRKISDKGGFNWELRVNVSVDNYLIYGLHIDNTRD 472
Db 649 TWFDDIFNAKNQIASITNNIENISNKID-----LNEFIISNEDSSKEL-LDEIRK 698
QY 473 FNIKLVKGDIDMGKMDYKANGPDPKVTMDGNV-YLQTYGSDLNKA-----V 522
Db 699 YKQQFD-KIKDAMNTEVKSFE-NTLQKOIDSIKSNINELTNAYDIINTKANLDDKLNNY 756
QY 523 GVHYQFLYDNVXP-----RVNIDPKGNTSIEYAD-GKSVFVFNINDKRNNGPGBEIQEHI 576
Db 757 GSEFNKLVNAGSLDITQKNDEKVKQLNLEYLKNKQSIIEINDIVNN----- 805
QY 577 YINGKEYTSFNDIKQIDKTLNLIKIVKDFARNTTVKEFILNKDTGEVSE---LKPFRVT 633
Db 806 FI--KELIKFNTE--TNKSLN-ELLTND--DINDKFLYKELNKISTNNLLKIYK-- 855
QY 634 VTIQNGKMSSTIVSEEDPILPVYKGELEKGYQFDGWEISGPEGKKGAGYVINSKOTF- 692
Db 856 NEIDNVNEKLSIVIEINLOFINSLSTEFNQ-----SITSH-----INFLANTLA 900
QY 693 -LKPVPKIEEKEEENKPTFDVSKKONPVNHSQLNESHKEDLQREHSQKSDSTKD 751
Db 901 GINDVLNKLNLKIMADTTTRGDTNIR-----DEIKNQISSENINKSQKFNKNE--KD 950
QY 752 VTTATV-----LDKNMISSKSTNN 770
Db 951 LKKLISFNDKLNKYNISAGYTEYN 974
RESULT 4
Ti18444
hypotheical protein C0385c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: Ti18444
R:Lawson, D.; Bowman, S.; Barrell, B.
A:Reference number: 218935
A:Accession: Ti18444
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1650 <LAW>
A:Cross-References: UNIPROT:077328; UNIPARC:UPI00000748ED; EMBL:298547; NID:e1325376; PI
C:Genetics:
A:Map position: 3
A:Introns: 1597/3; 1625/3
A:Note: C0385c
Query Match 6.2%; Score 250; DB 2; Length 1650;
Best Local Similarity 20.8%; Pred. No. 0.003;
Matches 205; Conservative 134; Mismatches 320; Indels 328; Gaps 48;
QY 20 SLKDDTTGVHHQHNEBSIKEKSTFTIDRNISTIRDPENKDLKLIKXKPRE---VDDF 76
Db 394 SLNENYNERKIYKEINKEYSNKEYV-----HFKND-DSSIKKONSSCLDEQ 443
QY 77 TSETGR--MEEDYKYDDKGNIIAY--DDGTDLE---YETKLEIKSKIYGVLSPSKD 129
Db 444 KKKTYKTIIEQRYNFNDRNNNAYIKDDTHKKEGYLLNMIQOSEYKKG--SNNKM 501

QY 130 GHFEILKGIS-----NVSQNAKVYGNKYSI--EIKATKYDFHSKWTTFDLYANINDV 182
Db 502 DEMEYLNQHTWDFNINENLNKIYF-DDYEGVDPEKKKKLDDHIYTOQKEYKNNINIL 560
QY 183 DGLAPAGMRULFVKNDOKKAEIKIRMEPIK-----ETKSEYPVYSSYGNVIELGEG-- 235
Db 561 K-----DHLLNDKETKEKQNEIEIEEEKKNKIEIEEEKKNKIEIEEEKKNKIEIEEEKK 614
QY 236 ---DLSKNKPNLTKMESGKIYSSEKQOYLKONI-----LJR 271
Db 615 KKIEIEEEKKKKI--EMEEKNKIDDEKKNYVANDKIISHIDNVNKNIKIDALLDHIIEKK 673
QY 272 KGYALKVTTYNPGKTD-----MLEGNYVSKEDIAK-----IQKANPNL-- 310
Db 674 KTGHKEINLYKEIKNEYQKMLNDENSINLEHEKKNYTHQVNNNLCDTKMDLOKENKILTN 733
QY 311 RALSETTIYADSRNVEDGRSTOSVMSALDGNRIYQVFTFKMDKGEADKQGNLVDSK 370
Db 734 NDKKKTFLLSKSKNI-----TSNVLSSKIPG-----TLSTKKNLNTIKTKD--VTD 779
QY 371 SSKLVLPF-----KDDKE-----YTGEDKF-NVEA-----IKEDGSMFLFIDTK 407
Db 780 NEKKYVHDHDKNDIIRKKEFINIYKGRNANYVEIGSEVCNNKVNKVGDDNKVMVENK 839
QY 408 P-----VNLSMDKNYF-----NPSKSNKIYVRNPEF 433
Db 840 QGDDNNMVENKQGDENNVIKNIERSSESPWFTHRRKNITSTDTCAKNEQIKYPHIL 899
QY 434 YLRKISDKGFGNWEIR--VNESVDNY-----LIYGLHIDN----- 469
Db 900 Y-SNKNEDEKKSIFLKNINENIKKYNKDEKISTLEKKVFKKDNVITNNDDEKHTSSK 958
QY 470 -TRDPNI-----KLN----- 478
Db 959 INDDFNITIDKQKGLNPNVDLARKIKNETKILEKKSHSKIQNNLMKKKTPPTNNK 1018
QY 479 -VKGDIDMDGMKDYKANGPDPKVTMDGNVYL-----OTG 513
Db 1019 GISSTSSSTKFNKFCGIIIEKNKLSNLKYTCIRKNMNVDSIKLNDKADLYDKDKKTS 1078
QY 514 YSDLNKAVGVHYQFLYDNVNVKPEVNDIPKGNSTIEYADGKSVFN-----INDKRNNGPFG 569
Db 1079 FNDINRAAKGMNFK-----KRDV---PNKNMVDVTNKGKR-VFNPVTLNNYRN-- 1124
QY 570 EIQQHLYINGKEYTSFNDIKQIDKTLNLIKIVKDFARNTTVKEFILNKDTGEVSELKP 629
Db 1125 -----YIRSNK-----NNVK-----NGKQV-----GIKKIVLLKE--KQKSLHP 1156
QY 630 HRVTVTIQNGKEMSSTIVSEEDFILPVYKGELEKGYQFDGWEISGFEKGDAGYVINSLSK 689
Db 1157 EGVEADKKLSNRYNDKYLIEKDGFKDIINEEMEK-----YKNNK-MKYKI--K 1201
QY 690 DTFIKVPFKTIEEKEEENKPTFDVSKKONPVNHSQLNESHKEDLQREH---HSQKS 746
Db 1202 SNSIPIIKKIERKSNNDN---DNIKNDNDINSNNINSNNDKCLFLSEKDRVHLKON 1258
QY 747 DSTKDVATVLDKKNISKSTNNPNK 773
Db 1259 NIIVNNTMMFRKQSCNDNNTTSLKNK 1285

RESULT 5

Ti18427

hypotheical protein C0335c - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: Ti18427

R:Lawson, D.; Bowman, S.; Barrell, B.

A:Reference number: 218935

A:Accession: Ti18427

A:Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA	
A;Residues: 1-3724 <LAW>	
A;Cross-references: UNIPROT:O77320; UNIPARC:UPI000007935D; EMBL:Z98547; NID:61325376; PI	
C;Genetics:	
A;Introns: 307/1; 1545/2	
A;Note: C0335C	
Qy	5 IAESKFNGLNGKEGS--LKK-----DTTGVEHHHQENESIK-----EKSSFT 46
Db	249 IPKNLFENLNNKOHNDYLRNIILMDVNDINLPHEQEQNSLNKKNKLCITGTNKKBEKYM 308
Qy	47 I-----DANISTIRDFENKDLKLIKKEFREVDFTSTGRMEEDYKYDDKG 95
Db	309 IPKGTEYQDKEKESILTINQDKKKYKK--KKYSELDQSDNISSNNTLSKRYTYT--- 363
Qy	96 NIIAYDGTDLLEYTEKLEDEIKSIYGVLSPSKDGHEILGKISNVSKNAKVYVGNKYK 155
Db	364 -----CGMDKETKETDEQNK-----RNTEVASICNVCNVEKQKRTKNKK 407
Qy	156 IEIKATKYDFHSKTMTFDLYANINDIVDGLAFAGDMR--LFVKNDQKAEIKRMP-EK 212
Db	408 ---SVQKDDLVEENALID---NL-IIIDGINFDDVTREKIIDNNENIDNNIYVENNK 460
Qy	213 IKETKSYPYVSSYG-NVI-----ELGEGDLSKNKPDNLTQMESGK 252
Db	461 LKQDKS-YDLFSSGKNLILGVNEGEFNEEVFENIEKLEQERKDNKDK-----T 512
Qy	253 IYDSEKQOYLKNDIILRKGYALKVTTYNPGKTDMLGNGVYSKEDIKIQKAN-PNLR 311
Db	513 IYNNNEQTDDLNRNI-----NKIESINNNNDN---NNNINNKKEFNKIRTEHILNKE 562
Qy	312 ALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFKMDKGEAIDKQGNLTDS 371
Db	563 SISXHIKGPSRDKR-----IKLUTYTNKNEDSTFELKLELEIITNN 604
Qy	372 SKLVLFQKDDKEYTGEDKFNVEAIKEDGSMLEFIDTKPVLNLSMDKNYPNPSKSN----- 424
Db	605 KVVYV--EEDIIGSNEDEYHVLKE-----NLKEDANEYNDKENKQNTKE 650
Qy	425 ----KIYVRN-----PEFYLRGKISDKGFNWEILRVN---ESVVDNLYIYGLDLHIDNRD 472
Db	651 ILKSKVYLENKRTLEELKLRGK---NNIFKKDEKYNLSGVEIINEIQINEENKINDIQD 707
Qy	473 FNIK-----LNVKD--GDIMDGMKDYKANGFPDKVTDMD----- 505
Db	708 GNISKQKIIQSSRTNDTFNKIDSLNDDLEKEKKRKKSHQHFIDNLVAKDQNEISENINK 767
Qy	506 -----GNVYLQT-----GYSDLNKAVGVHY----- 526
Db	768 ICDNINNIIYDESINNIYDESINNIYDESINNIYDENINNIIYDENINNIIYDE 827
Qy	527 -----QFLYDNNVKPEVND-----PKGNTSTIE-----Y 549
Db	828 NINNIIYDEGINIKICDDNILENKIKTTNDIVQVVEENNESIEKELMISLANKDINNNTYMF 887
Qy	550 ADGKSVFVN-----INDKRNNGFGCEIOE----- 573
Db	888 KENVDFIFINKIKRESLLKIDKNIKDNNNDDEIYIMDNFVENDFIIHNKWEITNKELDPLEI 947
Qy	574 -----QHIYINGKEYTS---FNDIKQIIDKTLNIKIVKDFARNTTTVKEFILNKDTGE 623
Db	948 NTQNEFIENLQI KKKKYTNDHFNDADQMFYE--NN-KILNKDMKKKEQEFKTDFTGS 1005
Qy	624 VSELK-----PHRVTVTTQNGKEMSGSTIVSE 649
Db	1006 LQSHKIKKYNKGEKHKDKNNEEKNILYDENQVYSVLSDHKIEQDQIHISIQTNICDE 1065
Qy	650 EDFILPVYKGELEKYQPDGWEISFGEGKDGAGVYNLSKDTFTFKPVFKLEEKKEENK 709

Db	1066 NN-IEQINENSKKGYRISGTDN---ENKND-----MENKNDMEKK 1102
Qy	710 PTFDVSKKDNPQVNH-SQLNESHKXEDLQREHSHQSDSTKDVTAIVLDDKNNSKSTT 768
Db	1103 N--DMEKNDIEKNDMEKKNDMEKKNDMEKKNDME-----NENNENKSDI 1153
Qy	769 NNPKN 773
Db	1154 ENENK 1158
RESULT 6	
T28677	
Rhoptry protein - Plasmodium yoelii	
C;Species: Plasmodium yoelii	
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004	
C;Accession: T28677; C45521	
R;Keen, J.; Sinha, K.; Brown, K.; Holder, A.	
Mol. Biochem. Parasitol. 65, 171-177, 1994	
A;Title: A gene coding for a high molecular mass rhoptry protein of Plasmodium yoelii.	
A;Reference number: Z20508; MUID:95021522; PMID:7935623	
A;Accession: T28677	
A;Status: preliminary; translated from GB/EMBL/DBJ	
A;Molecule type: DNA	
A;Residues: 1-2269 <KE>	
A;Cross-references: UNIPROT:Q26223; UNIPARC:UPI0000080871; EMBL:L27838; NID:9457145; PID	
R;Keen, J.; Holder, A.; Playfair, J.; Lockyer, M.; Lewis, A.	
Mol. Biochem. Parasitol. 42, 241-246, 1990	
A;Title: Identification of the gene for a Plasmodium yoelii rhoptry protein. Multiple co	
A;Reference number: A45521; MUID:91101660; PMID:2270106	
A;Accession: C45521	
A;Status: preliminary	
A;Molecule type: DNA	
A;Residues: 2131-2269 <KE2>	
A;Cross-references: UNIPARC:UPI000017B646; GB:M34283	
Query Match 5.8%; Score 231.5; DB 2; Length 2269;	
Best Local Similarity 19.7%; Pred. No. 0.028;	
Matches 181; Conservative 147; Mismatches 330; Indels 259; Gaps 37;	
Qy	37 ESIEKSSFTIDRNISTIRDFENK-DLKKLIKKEFREVDFTSETGRMEEY----- 87
Db	109 EGLKLELKKIKDIIAKIEYVKNVTELKKEIKNNAYIDELANQSPYKVTGYIENKNTIY 168
Qy	88 -----DYKDDKNIIAYDDGTDLLEYTEKLEDEIKSIYGVLSPSKDGCH 132
Db	169 NTIKSYFDQIYEGDIDTFYNELSSIVKEDPIDIEDKT-KLENLRSKIDNVYKIQKWEI 227
Qy	133 E-ILGKISNVSKNAKVYVGNKYKSIKATPKYDFHSHKTMTFDLYANINDIVDGLAFAGDM 191
Db	228 ETVKSHLNNIETNNKL-----PNTILIKKIYDEISK-----ELNKMLEDFNK---EKEL 276
Qy	192 RLFEVKNDQKK---AEIKRMP-----KIKETSEYPYVSSYGNVIELGE 234
Db	277 SNKISDYDKKREQLESEYKSKMLEIRUHYNSQTVNNTKKEEAKQNYDKSNEHMTTPTNE 336
Qy	235 GDLSK-----NKPDLNLTQNES-----GKIYSD---SEKQOYL-LKQNI---ILRKGYAL 276
Db	337 DEISKIISSEYKTKMKDEILSKVNTYIDFPNKKYKETVNSEHSQFTELTDKIKAEVSDKELKK 396
Qy	277 KVTYTPGKTDMLGNGVYSKE--DIAKIQKANPNLJALSET-----TIYAD-- 321
Db	397 CEQSFNDKSLINETKNSIEKQYQNTLKKVDEYIKVCKSTKESITKPSKQTLKXMDL 456
Qy	322 SRNVEDGRSTQSVLMSALDGFNIIRYQVFT-----FQMN-----DKGEAI---- 361
Db	457 NQNIKTVKETNSIDKSVIEKFE---QILTQKQTKLENKKTFTFSLNNHEANNNELIKYFS 512
Qy	362 DKQGNLVTDSKSLVLFQKDDKEYTGEDKFNVEAIKEDGSMLEFIDTKPVLNLSMDKNYFNS 421
Db	513 DLKANLGINENMLYNOQFTEKEKTFND-----IKEQNIHINEEISKIKIHASINIS 566
Qy	422 KSNKIYVRNPEFYLRGKISDKGGFNVELRVNESVVDNLYIYGLDLHIDNTRDFNKLNVK- 480

[illegible]

Qy	152	NYKS -TEIKATKYVDFHSKTNWTFDLYAN-----INDI---VDGLAPAGDMRLVFKVNDQOKKA	203
Db	743	KIKNDIDTLNQKIDKISGIEILT-EIKKNSENHIDEIKGQIDKLKKVPNTKTF--NEDPKKEI	799
Qy	204	EIKI-RMPEKIKETKSEYPPVSSYGVNVLGEGD---LSKNKPDNLTKMES-GKIY-----254	
Db	800	EKKIENIVEKIDRKKNYKEIDKLNLNEISKIENDKTSLEKLNKINUSYGSKSLGNLFQOI	859
Qy	255	-SDSEKQOYLK-----DNI-----ILRKYA	275
Db	860	DEEKKAEHTIKAMEAYIDDLNIIKKKSQIEKEMNINMDIKWDIHKEMKALNISHDDYK	919
Qy	276	LKVT--YNPCKTMDLNGVYSEKEDIAKQKANPNLRALSETTIYADSRNVEDGRSTQS	333
Db	920	IYHTTSKNEEKISDIRKNSLKIQDFSESYINDIKKELEKNVLESQNNNTDINQ-----975	
Qy	334	VLSALDG-FNIIRYQVFTEKMDKGEADKQGNLAVTDSKSLVLFQKDKKEYTGE-----387	
Db	976	-YLSKIENIYNIL-----KLNKIKKIDK-----VKEYTDETEKNN	1010
Qy	388	DFENVE-----AIKEDGSLFDITKPNVLSMDKNYFNPSKNIYVRNPEFYLRGK	438
Db	1011	KKINAELSNSEKIIITQKENSSLKECOSK-IKSTIDNNYVSECIKN---ITNLKTYI---1063	
Qy	439	ISKGGFENWELRVNESVDNY-LIYGDLHIDNTRDENIKLVNKDGDIMDWGMKDYKANGF	497
Db	1064	VNEKNINTYFKNAEYQNVLNFFNNIEMADTKSQVI-LNIK-----KNGG-1109	
Qy	498	PDKVTDMDGNVLYOTGYSDLNAAVGHVQFLYDNVKNPEVNIDPKNGTNSIEYADGSKSVF	557
Db	1110	-----TNTDYNIELKEHKK-----KSNYYKDEAGKNTQEIKNKEL--1147	
Qy	558	NINDKRNNGFDGBIQEIGHYINGKEYTSPNDIKQIIDKTLNI-KIVVKDF--ARNTTVKE	614
Db	1148	-----PEKYEQVTVLLN-KYAV--ELKNKFDKTKYSEQIIEIKDKAHNTFTSQ	1195
Qy	615	FIANKDTGEVSELKPHRVTVTIQNGKMGSTIVSEEDFILPVYKGELEKGYQFDGWEISG	674
Db	1196	--ADKSEKKWNEIKNQIRIIDEVAKNNKSN-----KAILD-----1229	
Qy	675	PEGKDGAGYVINLSKDTFKIPVKPIEB---KKEBENKPTFDVSKKKNDPQV-----723	
Db	1230	-----IQLSVFPF-KIKFLKIKDLRTKSDCLKETKOTETKISNLSDIQETKLI	1278
Qy	724	NHSQLSNEHSRKEDLOREHSQKSDSTKDVATVLDKNNISSKSTNNPNK	773
Db	1279	ENKNILNTLEKLESUKNQKNIEDQKKE-----LDEVSKIKNIESNVNQ	1324
RESULT 8			
T18440			
hypothetical protein C0425w - malaria parasite (Plasmodium falciparum)			
C/Species: Plasmodium falciparum			
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004			
C/Accession: T18440			
R/Lawson, D.; Bowman, S.; Barrell, B.			
submitted to the EMBL Data Library, August 1997			
A/Reference number: Z18935			
A/Accession: T18440			
A/Status: preliminary; translated from GB/EMBL/DBJ			
A/Molecule type: DNA			
A/Residues: 1-4550 <LAW>			
A/Cross-references: UNIPROT:O77336; UNIPARC:UPI0000110116; EMBL:Z98547; NID:el3			
C/Genetics:			
A/Map position: 3			
A/Note: C0425w			
Query Match			
Best Local Similarity 5.7%; Score 228; DB 2; Length 4550;			
Matches 193; Conservative 119; Mismatches 347; Indels 236; Gaps 43;			
Qy	5	IARSKPKNLNGKKE---GSLKDDTTGV-----EHHHQENEESIKSSFTIDRNISTIRD	56

Query Match 5.7%; Score 228; DB 2; Length 4550;
Best Local Similarity 21.6%; Pred. No. 0.098;
Matches 193; Conservative 119; Mismatches 347; Indels 236; Gaps 43;

Qy 5 IAESKFNGLNGKE---GSLKKDTTGv-----BHHHQENESIKESFTIDRNISTIRD 56
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

Db 746 IEBSRSISQGNHNEIYGASIKCAQGVDRDNKL DGNHNDKKDDNKDGDG---DKKDDNKD 802

Qy 57 FENKDLKLLIKKKPREVDDFTSETGKMEBYDYKIDKGNIIAYDDGTDLEYE----- 109

Db 803 GDKDDDKGDKDYDDGD-----KYDDDDNKYDDDDI--YDDDDDFNFDHSETSK 853

Qy 110 -----TEKLDEIK-SKIYGLVSPSKDGHFELGKIS-NVSKNAKYVG 150

Db 854 RLSNYDLLKKKKGSLKXEDTLKCSKMINVQBEKSKSKYCMNTWNCNVKE-KNYFV 912

Qy 151 NNYKSIKATKYDFHSKTMFTFDLIYANINDIVDGLAFAGD-----MRLF-----VKON 198

Db 913 NDYMHIIINLSDHEMNINMWSKPGYNLFXDVEHFIHKRDFINLILKLIKELIIVQLQN 972

Qy 199 DQKAEIKIRMPKIKETKSEYFVSSYGNV-----IELGEGDLS-----KN 240

Db 973 NYKQVCVFCCKTTGK-TKCMFPSCSTYFHIYCYNKYMOHVRRKKNDLQCASQKSRKN 1031

Qy 241 KPDNLTKWE--SGKIYSDSEKQVLLKDNILIRKGYALVTTYNPGKTDMLGNGVYSKE 298

Db 1032 QRAASVRRGAEGKDKGGRSLDRKKDKVTRRSNKYEESANDSGKKN--NGDDKKKG 1089

Qy 299 DIAKIQKANPLRALSETTIYADSRNVEDGRSTOSVLSALDGFNIIRYQVTFKMDKG 358

Db 1090 NIGKNNNNNNDDSDNNNGDDSDNNNG-----DGNNGINGDNNNINGDGN 1143

Qy 359 EAIKDGNLVT-----DSSKLVLFSGDKKEYTGDKFNV 392

Db 1144 NNINGDGNNDNCLCEKNQDGNQKRRKRNKRNNDDTSSLVLLNKDDSRQKCKNYG- 1202

Qy 393 EAIKEDGSMFLPDTKPVNLSMDKNYFNPSKSKIYVNPPEFYLRGKI-SDKGGFNWELRV 451

Db 1203 -----KDHS-----DQEKDSPSKINNRSKN-----KVKDKVEGN-----I 1234

Qy 452 NESVVDNLYIGDLHIDTRDFNKLNVXGDG----IMDMGMKYKANGFPDKVTDMDGN 507

Db 1235 NDDKINN-----NTDDNKNNI--NGDNKNNINGDNKNNING--DKI---NNN 1278

Qy 508 VYLTQGYSDLNKAVGVHY-QFLYDVKVPEVNDIPKNTSIEYADGSKSVFNNDK----- 562

Db 1279 INGDKINNINGDKISKNYHSHYDN--EETHEQKNFN-KNKNTYRKRSINDKEMND 1335

Qy 563 ----RNGFDGE-----IOEOHIYINGKEVTS-----FNDIKOIIDKTLNLIKVVKDPA 607

Db 1336 ILNNKNDTDKLLKNYFLMEYQKIISSDKITSGISNNMKDINKO-----IKDIK 1387

Qy 608 RNTTVKGFILNKDTGEVSELKPHRTVTVTIQNGKEMSTIVSEEDFILPVYKGELEKYQF 667

Db 1388 DIKNIKI---KDIKIDIKIDIKIKNVNIKD-----IKSNED----- 1424

Qy 668 DGWEISGFEGKKDAGYV-----INLSKDT--FKIPVFKIIE--KKEENKPTFDVSKKK 718

Db 1425 ----IKDIKDVNDVSSVHLANKVNNKSDTQDFCPSDDKNINIIKKNKINIKTDQVY--- 1477

Qy 719 DNPVNHSQLNESHKEDLQREHHSQKSD-STKDVTATVLDKNNISSKSTTNPN 772

Db 1478 -NPLVDSSTDCNKNYKKEKAVKXKIGWIKNIKULHNNNNNNNNNNNNNNNN 1531

RESULT 9

T18429

hypoetical protein C0345w - malaria parasite (Plasmodium falciparum)

C/Species: Plasmodium falciparum

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C/Accession: T18429

R/Lawson, D.; Bowman, S.; Barrell, B.

submitted to the EMBL Data Library, August 1997

A/Reference number: Z18935

A/Accession: T18429

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1711 <LAW>

A/Cross-references: UNIPROT:O77322; UNIPARC:UPI000007873B; EMBL:Z98547; NID:e1325376; PJ

C:Genetics:

A;Note: PFC0345w

Query Match 5.5%; Score 221.5; DB 2; Length 1711;

Best Local Similarity 19.7%; Pred. No. 0.053;

Matches 160; Conservative 134; Mismatches 274; Indels 243; Gaps 37;

Qy 29 EHHHQENESIKKSS-----FTIDRNISTIROFENKDLK-----KLIKKKPREVD 74

Db 984 DNNDDNNMMYDSSSSNNHYILNDKRLN-MDNFINNNLNNSQNKVIEKNLEYIN 1042

Qy 75 DF-TSETGKMEBYDYKIDKGNIIAYDDGTDLEYETEKLDKISKIYGLVSPSKDGHFE 133

Db 1043 NVKLTITS-----NYEQSNNTNSKDE-----HNISDSKSKEDTLNLSRKSSE 1086

Qy 134 ILGKI--SNVSKNAKYVYGNKYKSIKATKYDFHSKTMFTFDLIYANINDIVDGLAFAGDM 191

Db 1087 YNNKILQSTNSKSLANGAYENNLPSGKKKNK-----GTVLKDI-EHINDIQD--KYPEDL 1138

Qy 192 RL-----FYKDNQDKAEIKIRMPKIKETKSEYFVSS-----YGNVIELGEGDLSKNKPD 243

Db 1139 NINCNVKYYIENBEKHL-----LPLEL-----EYNLVSSDEKFG-----LNKIKND 1179

Qy 244 NLTKMESGKIYSDSEKQVLLKDNILIRKGYALVTTYNPGKTDMLGNGVYSKEDIK 303

Db 1180 N-----NIIYMKHQHNYLYDN---QKHILFDTKN-----VSI 1212

Qy 304 QKANPLRALSETTIYADSRNVEDGRSTOSVLSALDGFNIIRYQVTFKMDNK--GEAI 361

Db 1213 QR-NNNINSVIKTNHYEVEKNKQDN-----YDNFTCDKKKKIYYNII 1255

Qy 362 DKDGNLVTSSKLVLPFGDKKEYTG-----EDKENVEAIKEDGSMFLPDTKPVNL-- 411

Db 1256 NSDKDIYHNN--IITYKNEKEGIGHNLNRNDKDTNFEKLLKDGKVEFLDTPFKDSYID 1312

Qy 412 --SMDKNYFNPSKSKIYVNPPEFYLRGKISDKGFENWELRVNVEVDNLYIGDLHIDN 469

Db 1313 CHNKKENILNMTNKN-----EDHQIIDVADK-IFN---ETNMITMDNNKIYDDKNVHE 1362

Qy 470 TRDFNKLNVXGDGIMDMGMKYKANGFPDKVTDMD--GNVYLTQGYSDLNKAVGVHYQ 527

Db 1363 KKCTHNDVIHNNMDILSTSIKNEENLFDITYQKNRIGDIYNNR----- 1407

Qy 528 FLYDVKVPEVNDIPKNTSIEYADGSKSVFNNDKNNNG-----PDGEIOEOHIYINGKEY 583

Db 1408 ---INILQEDDDDDNHN-----NHNNNNNNNKLIILFEYTKNDQMLNHNKNL 1452

Qy 584 TSFNDIKQIIDKTLNLIKVVKDPARNTTVKFEFLNKDITGEVSELKPHRTVTVTIQNGKEMS 643

Db 1453 EGTEESDFIEKKNKIKIKNKNESYHKIDESLNSNEKNKVSL-----LINNNKDSS 1505

Qy 644 STIVSEEDFILPVYKGELEKYQDFGWEISGFEGKKDAGVIVNLSKDTFKIPVFKIIEK 703

Db 1506 SV-----DNNKNNNNNNNNNNNN-----NNNN 1527

Qy 704 KEENKPTFDVSKKDNQPNVHSQLNESHKED-----LQREHHSQKSDSTKDVTATV 756

Db 1528 NNNKNNNNNDSFKONNLINNNNNNNNNNDSFKONNLINNNNNNNNNNNNNNNNNKVIKEI 1587

Qy 757 LD---KNNTSSKST-----TNNPN 772

Db 1588 IDDKKNDIHKRDNIIYKOVSVSPLNNHPN 1618

RESULT 10

A24594

probable major surface antigen (83K, 19K, 42K) precursor - malaria parasite (Plasmodium

C/Species: Plasmodium falciparum

C/Date: 29-Aug-1987 #sequence_revision 29-Aug-1987 #text_change 31-Dec-2004

C/Accession: A24594

R/Holder, A.A.; Lockyer, M.J.; Odink, K.3.; Sandhu, J.S.; Riveros-Moreno, V.; Nicholls, J.

Nature 317, 270-273, 1985

A/Title: Primary structure of the precursor to the three major surface antigens of Plasmo-

Qy 551 DGKSVVNIN-----DKRNNGPDG---EIQEQHIYINGKEYTSFNDIQI IDKTL 597

Db 215 VLS--FNQEA VKDSFEFFYKEIVSKLKESKYNIDKKVDEITLTKTFADF-----SFEQ 267
QY 175 YANINDIVDGLAFAGDMRLFVKNDQKAEIKIRMEKIKETKSEYVYVSSYGNVIELGE 234
Db 268 Y--IEKIKDLFWAST--IKDTYQ-----TNKIFLNDVY----- 298
QY 235 GDSKKNKPNLWME-----SKYYS-----DSEKQYLLKDNILIRKGYAL 276
Db 299 --IKKNESNLKDKWFSNNFENEIYKFELEKDALVEIQALHQVLEKNEIITGEIVD 356
QY 277 KYTTPNGKTDLENGN-VYSKEDTAI-----QKANPNLRALS 314
Db 357 KVAAFIKSRELKSKAGLLFSKQDSAKITQLINHLISRYFPEAPENLELSKAKMGLNEL 416
QY 315 ETTIYADSRNVEDGRST-QSVLMSALDGFNIIRYQVTFKMDKGEAIDKGNLVTDSK 373
Db 417 KPEIYVEKEGEIAYPTLQDAISNAQDGQKIFL-----NKNLKLDK--SIVUDKN- 464
QY 374 LVLFGKDDKEYTGED--KPNVBAIKEDGSMFLFDTKPNVLSMDKNYFNPS----- 421
Db 465 ITIFAKSNVTITRKDSSKSFMTFIVQKGALTPEIAEPSSQSNLNGLGTSFKDESSLVKI 524
QY 422 -KSNKIYVNPFFYLRGKISDKGFENWELRVNESVVDNLYIYGDHLIDNTRDFN----- 474
Db 525 EKNALVAKTGTAFINSKSFYSKYG-----SVFEN---YGSVVIIEGAKIWNVSESG 572
QY 475 -----IKLVKDGDIDMDGMKDYKANGFPDK-VTDMGNNVLOTGYSDLNKAKAVGVH 525
Db 573 GIIRNHVGSLLTFKNGEI-----RDNISTG--DKGIISQGNIAISGSDGNKS----- 620
QY 526 YQLFDYDNKPEVNIIDPKNGTSTIEYADGKSVFNINDKR-----NNGFDGEIQEHYIING- 580
Db 621 FRSSLINLE-KTNINFNSGSI VNNASVKSILFEIDNSKIQISNNALINPFGSSAIFLKN 679
QY 581 -----KEYTSFNIDIKQIDKTLNIIKIV----- 602
Db 680 STMHLASLEKIKGEASEQRIEVLVDPQAKLISPKNIISLDNYQKLSSAIFKIFSVKN 739
QY 603 VKDFAR-----NTTWKEFLNKDT---GEVSELKPHRVTVTIQNGKMSSTIVSEE-DF 652
Db 740 INDFKHVPLVWNTKEKFFKLWPDTKLFVNIFYTKLQNHDLILQSGDFESTEKIIXDELD 799
QY 653 -----ILPVYKBLEGYQF-DGWEISGFEGK-----DAGYVINLSK 689
Db 800 YFRPTAAVKKLILTQVTRIPKYHKFEAFEPNFDLQKWYETINELIRLDPFYLFIAY 859
QY 690 DTFIKPVFKIEKEEENKPTFDVSKKDNQVNHSQLNSHREDLQR 739
Db 860 PEFV-----ENGMPLKPEY-IHTNVVNPVLEH-----FRNEDVAR 894
RESULT 14
B42771
reticulocyte-binding protein 2 - Plasmodium vivax (fragment)
C:Species: Plasmodium vivax
C:Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 09-Sep-1997
C:Accession: B42771
R:Galinski, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.
Cell 69, 1213-1226, 1992
A:Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites.
A:Reference number: A42771; PMID:92315338; PMID:1617731
A:Accession: B42771
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1252 <GAL>
A:Cross-references: UNIPARC:UPI000017B645; GB:M88098; NID:g160627; PID:g160628
A:Experimental source: strain Belem, merozoites
C:Genetics:
A:Gene: RBP2

Query Match

Best Local Similarity 5.4%; Score 215.5; DB 2; Length 1252;

Matches 207; Conservative 130; Mismatches 324; Indels 331; Gaps 50;

QY 2 LGEIAESKFK--NLNGKEGSLKDDTTGVEHHHOENESIEKESFT-IDRNI----- 51
Db 266 LSEIKYKDKCTTEISNSKRG---KDKTEFLKPKPNEESNSKNVINEINENIRNBOY 322
QY 52 -----STIRDFENKDLKLIK--FREVDFTS 78
Db 323 LKIDIEAQKASTKVELFHKHETTISNIFKSEILGVETKSQKINKAEIADIMKEIERHNS 382
QY 79 ETGKMEBYDYKDDKGNIIAYDDGTLEYTEKLEIKSIYGVLSKDGHPFLGKI 138
Db 383 EIQTOVRGFOENLKNLPHNYDNAED--ELANDKSTNAKVLITETLNVESVKN-----LSEI 437
QY 139 SNVSNAKVYCYNNYKSI--BIKATKYDFHSKTM-----TFDLYAN-INDIVIDGLAFAGD 190
Db 438 TNIOGGEKIY-SKAKDIMQIKATSENATAETLKVKKDDQSNVYNLNOIT--E 490
QY 191 MRLFYKND-----QKKAIEKIRMEPIKE-----TK-----S 218
Db 491 RNLIIVTEKNRLNGIDSTITNIEGALKESKGNVEIGFLEKLEBEIGKRNKLKVDITKKSINS 550
QY 219 EYPYVSSYGNVIELGEGDLKSNKPDNLTKMBGSKIYSDSEKQOYLLKKNILIRKGYALKV 278
Db 551 TVGNFSSLFNFDFLQYDFNKNINDYENKM--GBIYNEFEGSLNKISENL---RNASENT 605
QY 279 TTYNPGKTDMLEGNGVYSKEDI---AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVL 335
Db 606 SDYNSAKTLRLA---QKEKVNLLNKEEANKYLR-----DVKKVESFR----- 646
QY 336 MSALDGFNIIRYQVTFKQNDKGEAIDKGNLVTDSKLVFGKDDKEYTGBDKFNVEAI 395
Db 647 -----FIFNMK--ESLDKINEMIKKEQLTV-----NEGHGNVKQLVNI 683
QY 396 KEDGSMFLDTKPNLS-----MDKN-----YFNPSKSNKIYVRNPEFYLRGKI 439
Db 684 KE-----LVDEN--NLSDILKQATGKNBEIOKITHSTLNKAKTLGHVDTSAKYVGKI 736
QY 440 SDKGNNFELRVNWSVVD-----NLYIYGDHLIDNTRDFNKLNVKD--- 481
Db 737 TP-----ELALTTELLGDAKLTAQELKFESQNNVLETNMKNKTNLDVHKNIQDAYK 790
QY 482 -----GDIMDMGMKDYKANGFPDKVTDMDGNVYL-----QTGYSDLNKAKAVGVH 525
Db 791 VALEILAHSDSIDTKQD-----SSKLIEMGNQIYLVKVLINQYKNKISSIKSKEEAVS 844
QY 526 YQFLYDYNK-----PEVNIDPKNGTSTIEYADGKSVFNI-----NDKRNNGFDGEIQEQ 574
Db 845 VKI--GNVSKKHSBELSKITCSKSYDNIIALEKOTELQNLNRNSFTQKNTNDSKLE-- 900
QY 575 HIYINGKEYTSFNIDIKQIDKTLNIIKI-----VVKD--- 605
Db 901 -----KIKTDFESLKNAL-KTLEGEVNALKASDNHEHVQSKSEPPVNPALSEIEKETD 953
QY 606 -FARNTTVKEFILNKOTGEVSELKPHRVTVT--IQNGKEMSTIVSEEDFILPVYKGBLE 662
Db 954 IDSLNTALDELLKGRTCESRYKLIKDTVTKEISDDDELINTIEKN---VKAYLAVIK 1009
QY 663 KGYQPDGWEISGPGKKGAGVIVNLSKDTFLKPV-----FKKIEEKEEENKPTFD-- 713
Db 1010 KNYE-----DTVQD---VLTINHEFNTKQVSNHEPTNFDKSNKSEELTKAVTDSK 1057
QY 714 --VSKKK-----DNPOVN-----HSQLNESHR-KDQLQREHSQ 744
Db 1058 TIISKLKGVITVENMENTEMTIESAKEIEALYNELKNKKTSLNIEIYQTSNEVKLOEMKS 1117
QY 745 KSDSTKDVTA---TVLD---KNNISSKSTTN 770
Db 1118 NADKYVIDSVKIFNTVLTQKSNIVTNOHSINN 1149
RESULT 15
A45597
DNA-directed RNA polymerase (EC 2.7.7.6) III largest chain - malaria parasite (Plasmodium

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2006, 22:21:46 ; Search time 96 Seconds
(without alignments)
3681.545 Million cell updates/sec

Title: US-10-067-385-8

Perfect score: 4026

Sequence: 1 KLGEIAESKFNGLNGKEGS.....ATVLKNNISSKTTNNPNK 773

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_8.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*
- 9: Geneseqp2005s.*
- 10: Geneseqp2006s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4026	100.0	773	4 AAB48343	Aab48343 S. pneumo
2	4026	100.0	2140	6 ABU01020	Abu01020 S. pneumo
3	4026	100.0	2140	6 ABU45746	Abu45746 Protein e
4	4026	100.0	2140	8 ADM92113	Adm92113 S. pneumon
5	4026	100.0	2140	8 ADT50099	Adt50099 S. pneumon
6	4015.5	99.7	2120	3 AAY81710	Aay81710 Streptoco
7	3789.5	94.1	2138	8 ADK48759	Adk48759 Streptoco
8	3048	75.7	637	8 ADR94534	Adr94534 Novel S.
9	3048	75.7	637	9 AEA58404	Aea58404 Streptoco
10	750.5	18.6	1529	8 ADR96136	Adr96136 Novel S.
11	750.5	18.6	1529	9 AEA60006	Aea60006 Streptoco
12	615	15.3	117	2 AAW55096	Aaw55096 Streptoco
13	615	15.3	117	5 ABP54590	Abp54590 S. pneumo
14	615	15.3	117	7 ADC45149	Adc45149 S. pneumo
15	227	5.6	1196	6 ABU24813	Abu24813 Protein e
16	222	5.5	861	7 ABO23608	AbO23608 Plasmodiu
17	220	5.5	1166	8 ADK99186	Adk99186 Streptoco
18	218	5.4	1639	2 AAW54145	Aaw54145 P. falcip
19	218	5.4	1639	5 AAE29345	Aae29345 Plasmodiu
20	217	5.4	1166	8 ADK99185	Adk99185 Streptoco
21	217	5.4	1166	8 ADK99178	Adk99178 Streptoco
22	217	5.4	1166	8 ADK99184	Adk99184 Streptoco
23	217	5.4	1233	8 ADV82938	Adv82938 Streptoco

24	217	5.4	1239	8 ADV89524	Adv89524 Streptoco
25	217	5.4	1239	8 ADV80777	Adv80777 Streptoco
26	216	5.4	1166	8 ADK99183	Adk99183 Streptoco
27	216	5.4	1166	9 ADZ02952	Adz02952 S. agalac
28	216	5.4	1196	9 ADZ02951	Adz02951 S. agalac
29	216	5.4	1203	9 ADZ02950	Adz02950 S. agalac
30	216	5.4	1233	5 ABP30203	Abp30203 Streptoco
31	216	5.4	1233	5 ABP29675	Abp29675 Streptoco
32	216	5.4	1233	8 ADK99176	Adk99176 Streptoco
33	216	5.4	1233	8 ADO10469	Ado10469 Group B S
34	216	5.4	1233	9 ADZ02949	Adz02949 S. agalac
35	216	5.4	1239	5 ADP25822	Adp25822 Streptoco
36	215.5	5.4	1254	2 AAR07503	Aar07503 Merozoite
37	215.5	5.4	1254	2 AAW24575	Aaw24575 Merozoite
38	215	5.3	1166	8 ADK99179	Adk99179 Streptoco
39	212.5	5.3	2485	3 AAB18172	Aab18172 Plasmodiu
40	212	5.3	1141	6 ABU42327	Abu42327 Protein e
41	212	5.3	1188	3 AAB18183	Aab18183 Plasmodiu
42	211.5	5.3	991	8 ADP86452	Adp86452 Clostridi
43	211.5	5.3	1104	7 ADG73658	Adg73658 C. perfr
44	209	5.2	1558	3 AAB18324	Aab18324 Plasmodiu
45	208.5	5.2	3029	8 ADP25433	Adp25433 Plasmodiu

ALIGNMENTS

RESULT 1

AAB48343
ID AAB48343 standard; protein; 773 AA.
AC AAB48343;
XX
DT 20-APR-2001 (first entry)
XX
DE S. pneumoniae Spi130 polypeptide.
XX
KW Immunogenic; Spi128; Spi130; pneumococcal; otitis media; nasopharyngeal;
KW bronchial; lung; blood; infection; immune response; immunotherapy;
KW antibacterial; auditory; vaccine.
XX
OS Streptococcus pneumoniae.
XX
PN WO200076540-A2.
XX
PD 21-DEC-2000.
XX
PF 09-JUN-2000; 2000WO-US015925.
XX
PR 10-JUN-1999; 99US-0138453P.
XX
PI (MEDI-) MED IMMUNE INC.
XX
PI Adamou JE, Choi GH;
XX
DR WPI; 2001-112197/12.
XX
DR N-PSDB; AAC84742.
XX
PT New vaccines comprising Spi128 or Spi130 polypeptides, for treating and
PT preventing pneumococcal infections, particularly infections caused by
PT Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or
PT blood infections.
XX
PS Claim 8; Page 51-54; 54pp; English.
XX
CC The invention relates to novel immunogenic polypeptides, Spi128 and Spi130
CC from S. pneumoniae. Vaccines comprising the polypeptides are useful for
CC the treatment and prevention of pneumococcal infections, particularly
CC infections caused by Streptococcus, such as otitis media, nasopharyngeal,
CC bronchial, lung or blood infections. The antigens are used as immunogenic
CC agents to stimulate an immune response. The antisera and antibodies may
CC also be used in diagnosing and treating pneumococcal infections.
CC Recombinant polypeptides serve as a mechanism for stimulating production

CC of antibodies for use in passive immunotherapy, diagnostic reagents, and
CC as reagents in other processes such as affinity chromatography. The
CC present sequence represents the S. pneumoniae Sp130 polypeptide
XX
SQ Sequence 773 AA;

Query Match 100.0%; Score 4026; DB 4; Length 773;
Best Local Similarity 100.0%; Pred. No. 9e-225;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLGEIAESKPNLGNKGESLKKDTTGVVHHQNEESIEKSSFTIDRNIISTIRDPENK 60
Db 1 KLGEIAESKPNLGNKGESLKKDTTGVVHHQNEESIEKSSFTIDRNIISTIRDPENK 60
Qy 61 DLKKLIIKKKPREVDFTSETGKRMEEDYDYKDDKGNIIAYDDGTDLEYETEKLDEIKSKI 120
Db 61 DLKKLIIKKKPREVDFTSETGKRMEEDYDYKDDKGNIIAYDDGTDLEYETEKLDEIKSKI 120
Qy 121 YGVLSPSKDGHFELIGKISNVSKNAKYVYNNYKSIEIKATKYDFHSHKTMFTFDLIYANIND 180
Db 121 YGVLSPSKDGHFELIGKISNVSKNAKYVYNNYKSIEIKATKYDFHSHKTMFTFDLIYANIND 180
Qy 181 IVDGGLAFAGDMRLFKVNDQKAEIKIRMPKIKETKSEYVYSSYGNVIELGEGDLSKN 240
Db 181 IVDGGLAFAGDMRLFKVNDQKAEIKIRMPKIKETKSEYVYSSYGNVIELGEGDLSKN 240
Qy 241 KPDNLTKMESGKIYSDSEKQOYLLKDNIIILKGYALKVTTYNPGKTDMLEGNGVYSKEDI 300
Db 241 KPDNLTKMESGKIYSDSEKQOYLLKDNIIILKGYALKVTTYNPGKTDMLEGNGVYSKEDI 300
Qy 301 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIRIQVFTFKNDKGEA 360
Db 301 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIRIQVFTFKNDKGEA 360
Qy 361 IDKGNLVTDSKLVLFKGDKEVTGEDKFNVEAIKEDGSMFLFIDTKPVLNSMDKNVFN 420
Db 361 IDKGNLVTDSKLVLFKGDKEVTGEDKFNVEAIKEDGSMFLFIDTKPVLNSMDKNVFN 420
Qy 421 SKSNKIYVRNPEFYLRGKISDKGFNMLRVNESVDNLYIYDGLHIDNTRDFNKLNVK 480
Db 421 SKSNKIYVRNPEFYLRGKISDKGFNMLRVNESVDNLYIYDGLHIDNTRDFNKLNVK 480
Qy 481 DGDIMDWGMKDYKANGFPDPKVTMDGNYLQTYSDLNKAVGVHYQFLYDNVKEPVNID 540
Db 481 DGDIMDWGMKDYKANGFPDPKVTMDGNYLQTYSDLNKAVGVHYQFLYDNVKEPVNID 540
Qy 541 PKGNTSIEYADGKSVFVFNIDKRNNGPDGEIQEQHIIYINGKEYTSFNDIKQIIBDKTLNIK 600
Db 541 PKGNTSIEYADGKSVFVFNIDKRNNGPDGEIQEQHIIYINGKEYTSFNDIKQIIBDKTLNIK 600
Qy 601 IVVKDFARNTTVKGFILNKDTGEVSELKPHRVTVVTIQNGKEMSSTIVSEEDFILPVYKGE 660
Db 601 IVVKDFARNTTVKGFILNKDTGEVSELKPHRVTVVTIQNGKEMSSTIVSEEDFILPVYKGE 660
Qy 661 LEKGYPFGWISGPEGKADAGYVNLNSKDTFIKPVFKKIEEKEENKPTFDVSKKKDN 720
Db 661 LEKGYPFGWISGPEGKADAGYVNLNSKDTFIKPVFKKIEEKEENKPTFDVSKKKDN 720
Qy 721 PQVNHSQLNESHKREDLQREHSOKSDSTKDVATVLDKNNISKSTTNPNK 773
Db 721 PQVNHSQLNESHKREDLQREHSOKSDSTKDVATVLDKNNISKSTTNPNK 773

RESULT 2
ABU01020
ID ABU01020 standard; protein; 2140 AA.
XX ABU01020;
AC ABU01020;
XX
XX 23-OCT-2003 (revised)
DT 11-FEB-2003 (first entry)
XX
DE S. pneumoniae type 4 strain protein from coding region #590.

Query Match 100.0%; Score 4026; DB 6; Length 2140;
Best Local Similarity 100.0%; Pred. No. 3.5e-224;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLGEIAESKPNLGNKGESLKKDTTGVVHHQNEESIEKSSFTIDRNIISTIRDPENK 60
Db 1334 KLGEIAESKPNLGNKGESLKKDTTGVVHHQNEESIEKSSFTIDRNIISTIRDPENK 1393
Qy 61 DLKKLIIKKKPREVDFTSETGKRMEEDYDYKDDKGNIIAYDDGTDLEYETEKLDEIKSKI 120

XX
KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
KW gene therapy; vaccine.
XX
OS Streptococcus pneumoniae; type 4 strain.
XX
PN WO200277021-A2.
XX 03-OCT-2002.
XX 27-MAR-2002; 2002WO-IB002163.
XX 27-MAR-2001; 2001GB-00007658.
XX (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX Masignani V, Tettelin H, Fraser C;
XX WPI; 2003-040579/03.
XX N-PSDB; ABX06302.
XX
XX New proteins and nucleic acid molecules from Streptococcus pneumoniae.
XX useful as medicaments for treating or preventing a disease or infection
XX due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
XX ear infection.
XX
XX Claim 1; SEQ ID NO 1180; 56pp; English.
XX
CC The invention relates to a protein comprising or having at least 50%
CC identity to any of the 2469 amino acid sequences, identified in the
CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC ABS56454. Also included are an antibody which binds one of the proteins,
CC treating a patient by administering the protein, DNA or antibody (in a
CC composition), a kit comprising first and second primers, which are the
CC nucleic acid cited above or fragments between nucleotides 8-100 of a
CC sequence not defined in the specification, for amplifying a target
CC the first primer is substantially complementary to the target sequence
CC and the second primer is substantially complementary to the complement of
CC the target sequence, and where the parts of the primers having
CC substantial complementarity define the termini of the target sequence to
CC be amplified, assay comprising contacting a test compound with the
CC protein, and determining whether the test compound binds to the protein
CC and a Streptococcus pneumoniae bacterium, where one or more genes
CC encoding the proteins has been rendered inactive. The proteins, nucleic
CC acid molecules, antibody and compositions are useful as medicaments for
CC treating or preventing a disease or infection due to streptococcus
CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
CC media or ear infection. They are also useful in developing vaccines,
CC diagnostics and antibiotics. The methods are useful for identifying
CC immunodominant proteins. The present sequence is one of the 2469 proteins
CC expressed by the identified coding regions from the genomic sequence.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
CC standardise OS field)
XX
SQ Sequence 2140 AA;

Db	1394	DLKKLTKKKFREVDDTSETGKRMEEYDYKYDDKGNIIAYDDGTDLEVETEKLEDEIKSKI	1453
QY	121	YGVLSPSKDGHFELGKISNVSKNAKYVYGNKYSEIKATKYDFHSHKWTFTDLYANIND	180
Db	1454	YGVLSPSKDGHFELGKISNVSKNAKYVYGNKYSEIKATKYDFHSHKWTFTDLYANIND	1513
QY	181	IVDGLAFAGDMRLFVKNDQKAEIKIRMPKIKETKSEYPYVSSYGNVIELGEGDLSKN	240
Db	1514	IVDGLAFAGDMRLFVKNDQKAEIKIRMPKIKETKSEYPYVSSYGNVIELGEGDLSKN	1573
QY	241	KPDNLTKMESGKIYSDSEKQYLLKDNIIILRGYALKVTTYNPGKTDMLGEGVYSKEDI	300
Db	1574	KPDNLTKMESGKIYSDSEKQYLLKDNIIILRGYALKVTTYNPGKTDMLGEGVYSKEDI	1633
QY	301	AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFKQNDKGEA	360
Db	1634	AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFKQNDKGEA	1693
QY	361	IDKQGNLVTDSKLVLPFGKDDKEYTGEDKFNVEAIKEDGSMULFIDTKPVNLSMDKNYFNP	420
Db	1694	IDKQGNLVTDSKLVLPFGKDDKEYTGEDKFNVEAIKEDGSMULFIDTKPVNLSMDKNYFNP	1753
QY	421	SKSNKIYVRNPEFYLRGKISDKGFNWLVRNVSVDNYLIYDGLHIDNTRDFNIKLVK	480
Db	1754	SKSNKIYVRNPEFYLRGKISDKGFNWLVRNVSVDNYLIYDGLHIDNTRDFNIKLVK	1813
QY	481	DGDIWDGKMDYKANGFPDKVTDMDGNVYLQTYSDLNKAKAVGVHYOFLYDNNKPEVNI	540
Db	1814	DGDIWDGKMDYKANGFPDKVTDMDGNVYLQTYSDLNKAKAVGVHYOFLYDNNKPEVNI	1873
QY	541	PKGNTSIEYADGKSVFNNIDKRNNGFDGEIQEHYIYNGKEYTSFNDIKQIIDKTLNIK	600
Db	1874	PKGNTSIEYADGKSVFNNIDKRNNGFDGEIQEHYIYNGKEYTSFNDIKQIIDKTLNIK	1933
QY	601	IWKDFARNTTVKFEILNKDTGEVSELKPHRTVTIIONGKEMSSSTIVSEEDFILPVYKGE	660
Db	1934	IWKDFARNTTVKFEILNKDTGEVSELKPHRTVTIIONGKEMSSSTIVSEEDFILPVYKGE	1993
QY	661	LEKGQFDGWEISGFGKKDAGYVNLNSKDTFIKPVFKKIEKKKEENKPTFDVSKKKN	720
Db	1994	LEKGQFDGWEISGFGKKDAGYVNLNSKDTFIKPVFKKIEKKKEENKPTFDVSKKKN	2053
QY	721	PQVNHSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK	773
Db	2054	PQVNHSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK	2106
RESULT 3			
ID	ABU45746	standard; protein; 2140 AA.	
XX	XX	ABU45746;	
AC	ABU45746;		
XX	XX		
DT	19-JUN-2003	(first entry)	
XX	XX	Protein encoded by Prokaryotic essential gene #31273.	
DE	XX	Antisense; prokaryotic essential gene; cell proliferation; drug design.	
KW	XX	Streptococcus pneumoniae.	
OS	XX	WO200277183-A2.	
PN	XX	03-OCT-2002.	
PD	XX		
XX	XX	21-MAR-2002; 2002WO-US009107.	
PR	XX		
PR	21-MAR-2001; 2001US-00815242.		
PR	06-SEP-2001; 2001US-00948993.		
PR	25-OCT-2001; 2001US-0342923P.		
PR	08-FEB-2002; 2002US-00072851.		
PR	06-MAR-2002; 2002US-0362699P.		
XX	XX		

(ELIT-) ELITRA PHARM INC.

Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

WPI; 2003-029926/02.
N-PSDB; ACA49616.

New antisense nucleic acids, useful for identifying proteins or screening
for homologous nucleic acids required for cellular proliferation to
isolate candidate molecules for rational drug discovery programs.

Claim 25; SEQ ID NO 73670; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of
the 6213 antisense sequences given in the specification where expression
of the nucleic acid inhibits proliferation of a cell. Also included are:
(1) a vector comprising a promoter operably linked to the nucleic acid
encoding a polypeptide whose expression is inhibited by the antisense
nucleic acid; (2) a host cell containing the vector; (3) an isolated
polypeptide or its fragment whose expression is inhibited by the
antisense nucleic acid; (4) an antibody capable of specifically binding
the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
proliferation or the activity of a gene in an operon required for
proliferation; (7) identifying a compound that influences the activity of
the gene product or that has an activity against a biological pathway
required for proliferation, or that inhibits cellular proliferation; (8)
identifying a gene required for cellular proliferation or the biological
pathway in which a proliferation-required gene or its gene product lies
or a gene on which the test compound that inhibits proliferation of an
organism acts; (9) manufacturing an antibiotic; (10) profiling a
compound's activity; (11) a culture comprising strains in which the gene
product is overexpressed or underexpressed; (12) determining the extent
to which each of the strains is present in a culture or collection of
strains; or (13) identifying the target of a compound that inhibits the
proliferation of an organism. The antisense nucleic acids are useful for
identifying proteins or screening for homologous nucleic acids required
for cellular proliferation to isolate candidate molecules for rational
drug discovery programs, or for screening homologous nucleic acids
required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
K. pneumoniae or *P. aeruginosa*. The present sequence is encoded by one of
the target prokaryotic essential genes. Note: The sequence data for this
patent did not form part of the printed specification, but was obtained
in electronic format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences

Sequence 2140 AA;

Query Match 100.0%; Score 4026; DB 6; Length 2140;
Best Local Similarity 100.0%; Pred. No. 3.5e-224;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	KLGEIAESKPNLNGKESLKKDTTGVHHHQQNEESIKEKSFITDRNISTIRDFENK	60
Db	1334	KLGEIAESKPNLNGKESLKKDTTGVHHHQQNEESIKEKSFITDRNISTIRDFENK	1393
QY	61	DLKKLTKKKFREVDDTSETGKRMEEYDYKYDDKGNIIAYDDGTDLDLEYETEKLEDEIKSKI	120
Db	1394	DLKKLTKKKFREVDDTSETGKRMEEYDYKYDDKGNIIAYDDGTDLDLEYETEKLEDEIKSKI	1453
QY	121	YGVLSPSKDGHFELGKISNVSKNAKYVYGNKYSEIKATKYDFHSHKWTFTDLYANIND	180
Db	1454	YGVLSPSKDGHFELGKISNVSKNAKYVYGNKYSEIKATKYDFHSHKWTFTDLYANIND	1513
QY	181	IVDGLAFAGDMRLFVKNDQKAEIKIRMPKIKETKSEYPYVSSYGNVIELGEGDLSKN	240
Db	1514	IVDGLAFAGDMRLFVKNDQKAEIKIRMPKIKETKSEYPYVSSYGNVIELGEGDLSKN	1573
QY	241	KPDNLTKMESGKIYSDSEKQYLLKDNIIILRGYALKVTTYNPGKTDMLGEGVYSKEDI	300
Db	1574	KPDNLTKMESGKIYSDSEKQYLLKDNIIILRGYALKVTTYNPGKTDMLGEGVYSKEDI	1633
QY	301	AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFKQNDKGEA	360

Db 1634 AKIQKPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFKWNDKGEA 1693
Qy 361 IDKGNLVTSSKLVLFGKDKKEYTGEDKFNVEAIKEDGSMFLFDTKPVNLMSMDKNYFNP 420
Db 1694 IDKGNLVTSSKLVLFGKDKKEYTGEDKFNVEAIKEDGSMFLFDTKPVNLMSMDKNYFNP 1753
Qy 421 SKSNKIYVRNPEFYLRGKISDKGFGNWEELRVNESVDNYLYIGDLHIDNTRDFNKLNVK 480
Db 1754 SKSNKIYVRNPEFYLRGKISDKGFGNWEELRVNESVDNYLYIGDLHIDNTRDFNKLNVK 1813
Qy 481 DGDIMDMGMDKYKANGPDKVTDMGNYLQGYSDLNKAVGVHYOFLYDNVKEPVNID 540
Db 1814 DGDIMDMGMDKYKANGPDKVTDMGNYLQGYSDLNKAVGVHYOFLYDNVKEPVNID 1873
Qy 541 PKGNTSIEYADGKSVFNFINDKRNNGFDGETQEQHIIYINGKEYTSFNDIKQIIDKTLNIK 600
Db 1874 PKGNTSIEYADGKSVFNFINDKRNNGFDGETQEQHIIYINGKEYTSFNDIKQIIDKTLNIK 1933
Qy 601 IVVKDFARNTTVKPEFILNKDTGEVSELKPHRVTVTIQNGKEMSSTIYSEEDFILPVYKGE 660
Db 1934 IVVKDFARNTTVKPEFILNKDTGEVSELKPHRVTVTIQNGKEMSSTIYSEEDFILPVYKGE 1993
Qy 661 LEKGYPDQWEISGFEGKKDAGYVINLSKDTFIKPVFKKIEEKEEENKPTFDVSKKDN 720
Db 1994 LEKGYPDQWEISGFEGKKDAGYVINLSKDTFIKPVFKKIEEKEEENKPTFDVSKKDN 2053
Qy 721 PQVNHSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 773
Db 2054 PQVNHSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 2106

RESULT 4
ID ADM92113 standard; protein; 2140 AA.
XX ADM92113;
AC ADM92113;
XX 03-JUN-2004 (first entry)
DT S pneumoniae antigenic protein sequence SeqID310.
XX pneumoniae antigenic protein sequence SeqID310.
DE antibacterial; gene therapy; Streptococcus pneumoniae infection;
KW antigenic.
XX Streptococcus pneumoniae.
OS WO2004020609-A2.
PN 11-MAR-2004.
PD 02-SEP-2003; 2003WO-US027401.
XX 30-AUG-2002; 2002US-0407082P.
PR (TUFT) UNIV TUFTS.
XX Camilli A, Hava DL;
XX WPI; 2004-239189/22.
XX N-PSDB; ADM91876.
DR New Streptococcus pneumoniae nucleic acid molecules, useful for
XX diagnosing, treating and preventing active infections of Streptococcus pneumoniae.
PS Claim 27; SEQ ID NO 310; 123pp; English.

XX This invention relates to novel isolated Streptococcus pneumoniae nucleic acid molecules and the antigenic polypeptides encoded by them. The invention may be useful for the production of compounds with an antibacterial activity or for gene therapy. The nucleic acid molecules, compositions and methods disclosed are useful for treating Streptococcus

CC pneumoniae infection. The present sequence is that of an S pneumoniae
CC protein of the invention.
XX
SQ Sequence 2140 AA;
Query Match 100.0%; Score 4026; DB 8; Length 2140;
Best Local Similarity 100.0%; Pred. No. 3.5e-224; Indels 0; Gaps 0;
Matches 773; Conservative 0; Mismatches 0;
Qy 1 KLGEIAESKFNGLNGKEGSLKKDITGVVEHHHGHQNEESIKEKSSFTIDRNIISTIRDFENK 60
Db 1334 KLGEIAESKFNGLNGKEGSLKKDITGVVEHHHGHQNEESIKEKSSFTIDRNIISTIRDFENK 1393
Qy 61 DLKKLIKKEPREVDFTSETGKRMEEYDYKYDDKGNIIAYDDGTDLEETEKLDEIKSKI 120
Db 1394 DLKKLIKKEPREVDFTSETGKRMEEYDYKYDDKGNIIAYDDGTDLEETEKLDEIKSKI 1453
Qy 121 YGVLSPSKDGHFELGKISNVSKNAKYVYGNVYKSIIEIKATKYDFHKTWTFDLYANIND 180
Db 1454 YGVLSPSKDGHFELGKISNVSKNAKYVYGNVYKSIIEIKATKYDFHKTWTFDLYANIND 1513
Qy 181 IVDLAPAGDMRFLVKDNDQKAEIKIRMPKIKETKSEYYPVSSYGNVIELSGEDLSKN 240
Db 1514 IVDLAPAGDMRFLVKDNDQKAEIKIRMPKIKETKSEYYPVSSYGNVIELSGEDLSKN 1573
Qy 241 KPONLTKMESGKIYSDSEKQOYLLKUNIIIRKGYALKVTTYNPGKTDMLSGNGVYSKEDI 300
Db 1574 KPONLTKMESGKIYSDSEKQOYLLKUNIIIRKGYALKVTTYNPGKTDMLSGNGVYSKEDI 1633
Qy 301 AKTQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFKWNDKGEA 360
Db 1634 AKTQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFKWNDKGEA 1693
Qy 361 IDKGNLVTSSKLVLFGKDKKEYTGEDKFNVEAIKEDGSMFLFDTKPVNLMSMDKNYFNP 420
Db 1694 IDKGNLVTSSKLVLFGKDKKEYTGEDKFNVEAIKEDGSMFLFDTKPVNLMSMDKNYFNP 1753
Qy 421 SKSNKIYVRNPEFYLRGKISDKGFGNWEELRVNESVDNYLYIGDLHIDNTRDFNKLNVK 480
Db 1754 SKSNKIYVRNPEFYLRGKISDKGFGNWEELRVNESVDNYLYIGDLHIDNTRDFNKLNVK 1813
Qy 481 DGDIMDMGMDKYKANGPDKVTDMGNYLQGYSDLNKAVGVHYOFLYDNVKEPVNID 540
Db 1814 DGDIMDMGMDKYKANGPDKVTDMGNYLQGYSDLNKAVGVHYOFLYDNVKEPVNID 1873
Qy 541 PKGNTSIEYADGKSVFNFINDKRNNGFDGETQEQHIIYINGKEYTSFNDIKQIIDKTLNIK 600
Db 1874 PKGNTSIEYADGKSVFNFINDKRNNGFDGETQEQHIIYINGKEYTSFNDIKQIIDKTLNIK 1933
Qy 601 IVVKDFARNTTVKPEFILNKDTGEVSELKPHRVTVTIQNGKEMSSTIYSEEDFILPVYKGE 660
Db 1934 IVVKDFARNTTVKPEFILNKDTGEVSELKPHRVTVTIQNGKEMSSTIYSEEDFILPVYKGE 1993
Qy 661 LEKGYPDQWEISGFEGKKDAGYVINLSKDTFIKPVFKKIEEKEEENKPTFDVSKKDN 720
Db 1994 LEKGYPDQWEISGFEGKKDAGYVINLSKDTFIKPVFKKIEEKEEENKPTFDVSKKDN 2053
Qy 721 PQVNHSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 773
Db 2054 PQVNHSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 2106
RESULT 5
ADT50099
ID ADT50099 standard; protein; 2140 AA.
XX
XX AC ADT50099;
XX
XX 13-JAN-2005 (first entry)
DT S_pneumoniae hyperimmune serum reactive antigenic protein Seq 177.
XX hyperimmune serum reactive antigen; antibacterial; vaccine;
XX

KW bacterial infection; pharyngitis; otitis media; pneumonia; bacteraemia;
KW sepsis; meningitis.
OS Streptococcus pneumoniae TIGR4.
XX WO20040922209-A2.
XX 28-OCT-2004.
XX 15-APR-2004; 2004WO-EP003984.
XX 15-APR-2003; 2003EP-00450087.
XX (INTE-) INTERCELL AG.
PA Meinke A, Nagy E, Hanner M, Dewasthaly S, Stierschneider U;
PI WPI; 2004-758335/74.
DR N-PSDB; ADT49955.
XX New hyperimmune serum reactive antigens from Streptococcus pneumoniae,
PT and encoding nucleic acid molecules, useful for diagnosing, preventing or
PT treating S. pneumoniae infections.
XX Disclosure; SEQ ID NO 177; 191pp; English.

XX This invention relates to novel nucleic acids encoding hyperimmune serum
CC reactive antigens, or fragments derived thereof. Specifically, it refers
CC to antigens selected from peptides and serum reactive epitopes that can
CC be used in pharmaceutical compositions that exhibit antibacterial
CC activity. The present invention describes a composition (including the
CC nucleic acid molecule, hyperimmune serum-reactive antigen or antibody)
CC that is useful for manufacturing a medicament such as a vaccine, which
CC can be used to treat or prevent bacterial infections, particularly S.
CC pneumoniae infections that cause pharyngitis, otitis media, pneumonia,
CC bacteraemia sepsis and meningitis. The antigen or its fragment may also
CC be used for isolating, purifying and/ or identifying an interaction
CC partner of the hyperimmune serum reactive antigen, as well as for
CC manufacturing a functional nucleic acid selected from aptamers and
CC spiegelmers or for manufacturing a functional ribonucleic acid selected
CC from ribozymes, antisense nucleic acids and siRNA. This polypeptide
CC sequence is a Streptococcus pneumoniae hyperimmune serum reactive antigen
CC of the invention.

XX Sequence 2140 AA;

Query Match 100.0%; Score 4026; DB 8; Length 2140;
Best Local Similarity 100.0%; Pred. NO. 3.5e-224;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KLGEAIESKFNGLNGKESLKKDTTGVHHHQEENESIEKESFTIDRNIETIRDFENK 60
DB 1334 KLGEAIESKFNGLNGKESLKKDTTGVHHHQEENESIEKESFTIDRNIETIRDFENK 1393
QY 61 DLKLLIKKFRVDDFTSETGKMEYDYKDDKGNIIAYDDGTLEYETEKLEIKSKI 120
DB 1394 DLKLLIKKFRVDDFTSETGKMEYDYKDDKGNIIAYDDGTLEYETEKLEIKSKI 1453
QY 121 YGVLSFKDGHFEILGKISNVSKNAKYVGNKYKIEIKATKYDFHSTKMTFDDLYANIND 180
DB 1454 YGVLSFKDGHFEILGKISNVSKNAKYVGNKYKIEIKATKYDFHSTKMTFDDLYANIND 1513
QY 181 IVDGLAFAGDMRLFVKNDQKAEIKIRMPKIKETKSEYPVSVSSYGNVIEGEGDLKN 240
DB 1514 IVDGLAFAGDMRLFVKNDQKAEIKIRMPKIKETKSEYPVSVSSYGNVIEGEGDLKN 1573
QY 241 KPDNLTKMESGKIYSDSEKQYLLKNDNIILRGYALKVTTYNPGKTDMLGNGVYSKEDI 300
DB 1574 KPDNLTKMESGKIYSDSEKQYLLKNDNIILRGYALKVTTYNPGKTDMLGNGVYSKEDI 1633
QY 301 AKIQANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGNIIIRYQVFTPKNDKGEA 360
DB 1634 AKIQANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGNIIIRYQVFTPKNDKGEA 1693

QY 361 IDKGNLVTDSSKLVLFKDKKEYTGEDKFNVEAIKEDGSMLEFIDTKPVLNSMDKNYFNP 420
DB 1694 IDKGNLVTDSSKLVLFKDKKEYTGEDKFNVEAIKEDGSMLEFIDTKPVLNSMDKNYFNP 1753
QY 421 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVVDNLYIYGDLDHIDNTRDFNKLANKV 480
DB 1754 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVVDNLYIYGDLDHIDNTRDFNKLANKV 1813
QY 481 DGDIMDWGMKDYKANGFPDKVYTDMDGNVYLOTGYSDLNKAKAVGVHYQFLYDYNVKEVIND 540
DB 1814 DGDIMDWGMKDYKANGFPDKVYTDMDGNVYLOTGYSDLNKAKAVGVHYQFLYDYNVKEVIND 1873
QY 541 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIOEHYINGKEYTSFNDIKQIIDTLNLIK 600
DB 1874 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIOEHYINGKEYTSFNDIKQIIDTLNLIK 1933
QY 601 IVVKDFARNTTVKEPILNKOTGEVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGE 660
DB 1934 IVVKDFARNTTVKEPILNKOTGEVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGE 1993
QY 661 LEKGYQFDGWEISGPEGKKDAGYVNLKSDTFIKPFVKKIEKKEENKPTFDVSKKXDN 720
DB 1994 LEKGYQFDGWEISGPEGKKDAGYVNLKSDTFIKPFVKKIEKKEENKPTFDVSKKXDN 2053
QY 721 PQVNHSQLNESHRKEDLOREHSHQSDSTKDVATVLDKNNISSTNNPNK 773
DB 2054 PQVNHSQLNESHRKEDLOREHSHQSDSTKDVATVLDKNNISSTNNPNK 2106
ID AAY81710 standard; protein; 2120 AA.
XX AAY81710;
XX 02-JUN-2000 (first entry)
DT Streptococcus pneumoniae protein sequence ID3.
DE Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;
KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;
KW kidney disease; diabetes; immunosuppressive disorder; otitis media;
KW pneumococcal septicaemia; sinusitis; meningitis; therapy.
OS Streptococcus pneumoniae.
XX WO200006738-A2.
XX 10-FEB-2000.
XX 27-JUL-1999; 99WO-GB002452.
XX 27-JUL-1998; 98GB-00016336.
XX 19-MAR-1999; 99US-0125329P.
XX (MICR-) MICROBIAL TECHNIQS LTD.
XX Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;
PI WPI; 2000-195301/17.
XX N-PSDB; AAZ91806.
XX Streptococcal proteins and polynucleotides useful for diagnosis,
PT treatment and prophylaxis of bacterial infections.
XX Claim 2; Page 41-42; 76pp; English.
PS This sequence represents a Streptococcus pneumoniae protein of the
CC invention. The proteins (or their homologues, derivatives and/or
CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic
CC compositions comprising the proteins are useful as vaccines and also in
CC diagnostic assays. The sequences are useful for the detection or

CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested
CC with them. Agents capable of antagonising, inhibiting or interfering with
CC the function or expression of the protein or polypeptide are useful in
CC medical compositions in the treatment or prophylaxis of S. pneumoniae
CC infection. As the sequences can be used to treat S. pneumoniae infection,
CC they can be used to treat bacterial pneumonia, which has high rates in
CC young children, the elderly, and in patients with predisposing conditions
CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,
CC or with immunosuppressive disorders, especially AIDS. They can also be
CC used to treat pneumococcal septicaemia, otitis media, sinusitis, and
CC meningitis
XX
SQ Sequence 2120 AA;

Query Match 99.7%; Score 4015.5; DB 3; Length 2120;
Best Local Similarity 99.9%; Pred. No. 1.4e-223;
Matches 773; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 KLGEIAESKFNKLGNGEGSLKDDTTGVEHHQNEESIKEKSSFTIDRNISTIRDENK 60
Db 1313 KLGEIAESKFNKLGNGEGSLKDDTTGVEHHQNEESIKEKSSFTIDRNISTIRDENK 1372
Qy 61 DLKKLIKKKREVDFTSETGRMEEDYKYDDKGNIIAYDDGTDLEVEYETKLEIKSKI 120
Db 1373 DLKKLIKKKREVDFTSETGRMEEDYKYDDKGNIIAYDDGTDLEVEYETKLEIKSKI 1432
Qy 121 YGVLSPSKDGHFELGKISNVSKNAKYVYNNYKSIIEIKATKYDFHSKTMTFDLYANIND 180
Db 1433 YGVLSPSKDGHFELGKISNVSKNAKYVYNNYKSIIEIKATKYDFHSKTMTFDLYANIND 1492
Qy 181 IVDGLAFAGDNRFLVKDNDQKABIKIRMPKIKETKSEYVYSSYNNVIELGEGDLSKN 240
Db 1493 IVDGLAFAGDNRFLVKDNDQKABIKIRMPKIKETKSEYVYSSYNNVIELGEGDLSKN 1552
Qy 241 KPDNLTKMESKIIYSDSEKQYLLKDNILIRKGVALKVTTYNPCKTDMLENGVYSKEDI 300
Db 1553 KPDNLTKMESKIIYSDSEKQYLLKDNILIRKGVALKVTTYNPCKTDMLENGVYSKEDI 1612
Qy 301 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFQWMDKGEA 360
Db 1613 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFQWMDKGEA 1672
Qy 361 IDKGNLVTDSKLVLFKGDDEYTGEDKFNVEAIKEDGSMFLFDTDKPVLMSMDKNYFNP 420
Db 1673 IDKGNLVTDSKLVLFKGDDEYTGEDKFNVEAIKEDGSMFLFDTDKPVLMSMDKNYFNP 1732
Qy 421 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVVDNLIYDGLHIDNTR - DFNIKLVN 479
Db 1733 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVVDNLIYDGLHIDNTRDDFNILVN 1792
Qy 480 KGDGIMDMGMDKYKANGFPDPKVTMDGNVYLQTYGSDLNKAKAVGVHYQFLVDNVKPEVNI 539
Db 1793 KGDGIMDMGMDKYKANGFPDPKVTMDGNVYLQTYGSDLNKAKAVGVHYQFLVDNVKPEVNI 1852
Qy 540 DPKGNSTIYADGKSVFNINDKRNNGPDGIEQHIYINGKEVTSNDIKQIIDKTLNI 599
Db 1853 DPKGNSTIYADGKSVFNINDKRNNGPDGIEQHIYINGKEVTSNDIKQIIDKTLNI 1912
Qy 600 KIVVKDFARNVTVKFELNKDTGVESELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKG 659
Db 1913 KIVVKDFARNVTVKFELNKDTGVESELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKG 1972
Qy 660 ELEKGYQDGEWISGFEKGKAGVYVNLKDTFFIKPVFKKIEKKKEENKPTFDVSKKDD 719
Db 1973 ELEKGYQDGEWISGFEKGKAGVYVNLKDTFFIKPVFKKIEKKKEENKPTFDVSKKDD 2032
Qy 720 NPQVNHSQLNSHRKEDLOREHSQKSDSTKDVATATVLDKNNISKSSTTNPNK 773
Db 2033 NPQVNHSQLNSHRKEDLOREHSQKSDSTKDVATATVLDKNNISKSSTTNPNK 2086

RESULT 7
ADK48759

ID ADK48759 standard; protein; 2138 AA.
XX
AC ADK48759;
XX
DT 20-MAY-2004 (first entry)
XX
DE Streptococcus pneumoniae protein, Seq ID No 5274.
XX
KW Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
XX
OS Streptococcus pneumoniae.
XX
PN US6699703-B1.
XX
PD 02-MAR-2004.
XX
PF 26-MAY-2000; 2000US-00583110.
XX
PR 02-JUL-1997; 97US-0051553P.
PR 12-MAY-1998; 98US-0085131P.
PR 30-JUN-1998; 98US-00107433.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CB;
XX
DR WPI; 2004-212399/20.
DR N-PSDB; ADK46098.
XX
PT New nucleic acid molecules and polypeptides useful for diagnosing,
PT preventing and treating pathological conditions resulting from bacterial
PT infection, e.g. Streptococcus pneumoniae infection, and in drug
PT screening.
XX
PS Disclosure; SEQ ID NO 5274; 301pp; English.
XX
CC The invention relates to isolated Streptococcus pneumoniae nucleic acids
CC and polypeptides. The nucleic acids and proteins are useful for
CC diagnosing, preventing and treating pathological conditions resulting
CC from bacterial infection, such as S. pneumoniae infection. These may also
CC be used for drug screening procedures. The present sequence represents a
CC Streptococcus pneumoniae polypeptide of the invention. Note: The sequence
CC data for this patent did not appear in the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 2138 AA;

Query Match 94.1%; Score 3789.5; DB 8; Length 2138;
Best Local Similarity 94.8%; Pred. No. 1.8e-210;
Matches 733; Conservative 15; Mismatches 24; Indels 1; Gaps 1;
Qy 1 KLGEIAESKFNKLGNGEGSLKDDTTGVEHHQNEESIKEKSSFTIDRNISTIRDENK 60
Db 1333 KLGEIPESKFNKLVKQVDDSLNKETAENVNLLVDNQSIIEGKSLFNHKITISTIRDENK 1392
Qy 61 DLKKLIKKKREVDFTSETGRMEEDYKYDDKGNIIAYDDGTDLEVEYETKLEIKSKI 120
Db 1393 DLKKLIKKKREVDFTSETGRMEEDYKYDDKGNIIAYDDGTDLEVEYETKLEIKSKI 1451
Qy 121 YGVLSPSKDGHFELGKISNVSKNAKYVYNNYKSIIEIKATKYDFHSKTMTFDLYANIND 180
Db 1452 YGVLSPSKDGHFELGKISNVSKNAKYVYNNYKSIIEIKATKYDFHSKTMTFDLYANIND 1511
Qy 181 IVDGLAFAGDNRFLVKDNDQKABIKIRMPKIKETKSEYVYSSYNNVIELGEGDLSKN 240
Db 1512 IVDGLAFAGDNRFLVKDNDQKABIKIRMPKIKETKSEYVYSSYNNVIELGEGDLSKN 1571
Qy 241 KPDNLTKMESKIIYSDSEKQYLLKDNILIRKGVALKVTTYNPCKTDMLENGVYSKEDI 300
Db 1572 KPDNLTKMESKIIYSDSEKQYLLKDNILIRKGVALKVTTYNPCKTDMLENGVYSKEDI 1631
Qy 301 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFQWMDKGEA 360

Db 1632 AKIQKANPNLRLSETTIYADSRNVEDGRSTQAVLMSALDGFNIIRYQVFTFMNDKGEA 1691
QY 361 IDKDGNLVTDSSKLVLFQKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPNVLSMDKNYFNP 420
Db 1692 IDKDGNLVTDSSKLVLFQKDDKEYTGEDKSNVEAIKEDGSMFLFIDTKPNVLSMDKNYFNP 1751
QY 421 SKSNKIYVRNPFFYLRGKISDKGGFNWELRVNESVVDNLIYGDHLIDNTRDFNKLNVK 480
Db 1752 SKSNKIYVRNPFFYLRGKISDKGGFNWELRVNESVVDNLIYGDHLIDNTRDFNKLNVK 1811
QY 481 DGDIMWGKDYKANGFPDKVTDMDGNVYLOTGYSDLNAKAVGVHYQFLYDNVKEVNID 540
Db 1812 DGDIMWGKDYKANGFPDKVTDMDGNVYLOTGYSDLNAKAVGVHYQFLYDNVKEVNID 1871
QY 541 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIOEOHIYINGKEVTSFNDIKQIIDKTLNLIK 600
Db 1872 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIOEOHIYINGKEVTSFNDIKQIIDKTLNLIK 1931
QY 601 IVVKOFARNTTVKEFILNKDTEGEVSELKPHRVTVTTIQNGKEMSSITIVSEEDFILPVYKGE 660
Db 1932 IVVKOFARNTTVKEFILNKDTEGEVSELKPHRVTVTTIQNGKEMSSITIVSEEDFILPVYKGE 1991
QY 661 LEKGYQFDGWEISGPEKKDAGVYNLSKDTFIKPVFKKIEBKKEENKPTTFDVSKKKDN 720
Db 1992 LEKGYQFDGWEISGPEKKDAGVYNLSKDTFIKPVFKKIEBKKEENKPTTFDVSKKKDN 2051
QY 721 PQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 773
Db 2052 PQVNHSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 2104
RESULT 8
ADR94534
ID ADR94534 standard; protein; 637 AA.
XX ADR94534;
XX
DT 16-DEC-2004 (first entry)
XX
DE Novel S. pneumoniae protein sequence, SEQ ID 3169.
XX
KW Meningitis; bacteraemia; pneumonia; otitis media; vaccine;
XX bacterial infection.
XX
OS Streptococcus pneumoniae.
XX
PN US6800744-B1.
XX
PD 05-OCT-2004.
XX
XX 30-JUN-1998; 98US-00107433.
XX
XX
PR 02-JUL-1997; 97US-0051553P.
PR 12-MAY-1998; 98US-0085131P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
XX WPI; 2004-697205/68.
DR N-PSDB; ADR91931.
XX
XX New isolated nucleic acid encoding a Streptococcus pneumoniae
PT polypeptide, useful for diagnosing, preventing and/or treating
PT pathological conditions resulting from the bacterial infection.
XX
PS Disclosure; SEQ ID NO 3169; 151pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising a sequence
CC encoding a Streptococcus pneumoniae ADR91366polypeptide, or its
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,
CC ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,
CC

CC ADR96079) or any of the fully defined sequences appearing as ADR91705,
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide
CC sequences or at least 40, 60 or 300 consecutive nucleotides, which is
CC hybridisable under high stringency conditions to the nucleotide sequence.
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.
CC Also included are a recombinant expression vector comprising the isolated
CC nucleic acid cited above operably linked to a transcription regulatory
CC element, a cell comprising the recombinant expression vector and a probe
CC comprising at least 20 consecutive nucleotides of the nucleotide
CC sequences as cited above. The methods and compositions of the present
CC invention are useful for the diagnosis, prevention and/or treatment of
CC pathological conditions resulting from bacterial infection by
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and
CC otitis media. The present sequence is one of the 2603 disclosed S.
CC pneumoniae protein sequences. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.
XX
SQ Sequence 637 AA;
Query Match 75.7%; Score 3048; DB 8; Length 637;
Best Local Similarity 99.3%; Pred. No. 3.4e-168;
Matches 585; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 185 LAFAGDMRLFVKDNDQKAEIKIRMEKIKETKSEYVPYVSSYGNVIELGEGDLSKNKPN 244
Db 15 LAFAGDMRLFVKDNDQKAEIKIRMEKIKETKSEYVPYVSSYGNVIELGEGDLSKNKPN 74
QY 245 LTKMESGKIYSDSEKQYLLKDNIIILRGYALKVTTYNPNGKTDMLLEGNGVYSEKIDIAKTQ 304
Db 75 LTKMESGKIYSDSEKQYLLKDNIIILRGYALKVTTYNPNGKTDMLLEGNGVYSEKIDIAKTQ 134
QY 305 KANPNLRLSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFMNDKGEAIDKD 364
Db 135 KANPNLRLSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFMNDKGEAIDKD 194
QY 365 GNLVTDSSKLVLFQKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPNVLSMDKNYFNP 424
Db 195 GNLVTDSSKLVLFQKDDKEYTGEDKSNVEAIKEDGSMFLFIDTKPNVLSMDKNYFNP 254
QY 425 KIYVRNPFFYLRGKISDKGGFNWELRVNESVVDNLIYGDHLIDNTRDFNKLNVKDGDI 484
Db 255 KIYVRNPFFYLRGKISDKGGFNWELRVNESVVDNLIYGDHLIDNTRDFNKLNVKDGDI 314
QY 485 MDWGMKDYKANGFPDKVTDMDGNVYLOTGYSDLNAKAVGVHYQFLYDNVKEVNIDPKGN 544
Db 315 MDWGMKDYKANGFPDKVTDMDGNVYLOTGYSDLNAKAVGVHYQFLYDNVKEVNIDPKGN 374
QY 545 TSIEYADGKSVVFNINDKRNNGFDGEIOEOHIYINGKEVTSFNDIKQIIDKTLNLIK 604
Db 375 TSIEYADGKSVVFNINDKRNNGFDGEIOEOHIYINGKEVTSFNDIKQIIDKTLNLIK 434
QY 605 DFARNTTVKEFILNKDTEGEVSELKPHRVTVTTIQNGKEMSSITIVSEEDFILPVYKGELEK 664
Db 435 DFARNTTVKEFILNKDTEGEVSELKPHRVTVTTIQNGKEMSSITIVSEEDFILPVYKGELEK 494
QY 665 YQFDGWEISGPEKKDAGVYNLSKDTFIKPVFKKIEBKKEENKPTTFDVSKKKDNPOVN 724
Db 495 YQFDGWEISGPEKKDAGVYNLSKDTFIKPVFKKIEBKKEENKPTTFDVSKKKDNPOVN 554
QY 725 HSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 773
Db 555 HSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNPNK 603
RESULT 9
AEA58404
ID AEA58404 standard; protein; 637 AA.
XX AEA58404;
XX

CC The invention relates to an isolated nucleic acid comprising a sequence
CC encoding a Streptococcus pneumoniae ADR91366polypeptide, or its
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,
CC ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,
CC ADR96079) or any of the fully defined sequences appearing as ADR91705,
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is
CC hybridizable under high stringency conditions to the disclosed sequence.
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.
CC Also included are a recombinant expression vector comprising the isolated
CC nucleic acid cited above operably linked to a transcription regulatory
CC element, a cell comprising the recombinant expression vector and a probe
CC comprising at least 20 consecutive nucleotides of the nucleotide
CC sequences as cited above. The methods and compositions of the present
CC invention are useful for the diagnosis, prevention and/or treatment of
CC pathological conditions resulting from bacterial infection by
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and
CC otitis media. The present sequence is one of the 2603 disclosed S.
CC pneumoniae protein sequences. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.
XX
XX Sequence 1529 AA;

Query Match 18.6%; Score 750.5; DB 8; Length 1529;
Best Local Similarity 79.7%; Pred. No. 1.6e-34;
Matches 149; Conservative 14; Mismatches 23; Indels 1; Gaps 1;
QY 1 KLGEIAESKFNKNGKESLKDITGVVHHQNEESIEKESFTIDRNFENK 60
Db 1336 KLGEIPESKFNKKNVDDSLNKETAENVNLLVDNQSIGKSLFNIHKTISTIRDFENK 1395
QY 61 DLKLLIKKKFREVDDFTSETGRMEEDYKYDDKGNIIAYDDGTDLLEYETEKLDIKSKI 120
Db 1396 DLKLLIKKKYKQEDDFVN-GGTRTVVERDYKYDDKGNIIAYDDGTDLLEYETEKLDIKSKI 1454
QY 121 YGVLSPSKDGHPFELIGKISNVSKNAKYVGNKYKIEIKATKYDFHSTMTFDLYANIND 180
Db 1455 YGVLSPSKDGHPFELIGKISNVSKNAKYVGNKYKIEIKATKYDFHSTMTFDLYANIND 1514
QY 181 IVDGLAF 187
Db 1515 IVDGISF 1521

RESULT 11
AEA60006
ID AEA60006 standard; protein; 1529 AA.
XX AEA60006;
AC AEA60006;
XX 25-AUG-2005 (first entry)
XX Streptococcus pneumoniae ORF amino acid sequence SEQ ID NO:4771.
DE Streptococcus pneumoniae ORF amino acid sequence SEQ ID NO:4771.
KW bacterial infection; Streptococcus pneumoniae infection; antibacterial;
KW vaccine.
XX Streptococcus pneumoniae.
OS US2005136404-A1.
PN 23-JUN-2005.
PD 10-JUL-2003; 2003US-00617320.
PF 02-JUL-1997; 97US-0051553P.
PR 12-MAY-1998; 98US-0085131P.
PR 30-JUN-1998; 98US-00107433.
XX (DOUC/) DOUCETTE-STAMM L A.

(BUSH/) BUSH D.

Doucette-Stamm LA, Bush D;

WPI: 2005-477576/48.

N-PSDB; AEA57403.

New isolated nucleic acid molecules and encoded polypeptides useful for
diagnosing, preventing or treating bacterial infections, particularly
Streptococcus pneumoniae infection.

Claim 5; SEQ ID NO 4771; 144pp; English.

The invention relates to an isolated nucleic acid molecule for detecting,
preventing or treating pathological conditions resulting from bacterial
infection. The isolated nucleic acid comprises: (a) any of the 2603
nucleotide sequences of AEA55236 to AEA57838; (b) a nucleotide sequence
encoding a Streptococcus pneumoniae polypeptide comprising any of the
2603 amino acid sequences of AEA57839 to AEA60441; or (c) a nucleotide
sequence of at least 8 nucleotides in length, where the sequence is
hybridizable to a nucleic acid having any of the nucleotide sequences in
(a). Also described: (1) a recombinant expression vector comprising the
above nucleic acid operably linked to a transcription regulatory element;
(2) a cell comprising the recombinant expression vector; (3) producing an
S. pneumoniae polypeptide; (4) a probe comprising a nucleotide sequence
consisting of at least 8 nucleotides of any of AEA55236 to AEA57838; (5)
treating a subject for S. pneumoniae infection; (6) a recombinant or
substantially pure preparation of an S. pneumoniae polypeptide or its
fragment, where the polypeptide is selected from AEA57839 to AEA60441;
(7) a vaccine composition for preventing or treating an S. pneumoniae
infection, comprising an amount of the above nucleic acid or polypeptide;
(8) detecting the presence of a Streptococcus nucleic acid in a sample;
(9) a computer readable medium having recorded the nucleotide sequences
of AEA55236 to AEA57838; (10) a computer based system for identifying
fragments of the Streptococcus genome of commercial importance. The
composition and methods are useful for diagnosing, preventing or treating
bacterial infections, particularly S. pneumoniae infection. The present
sequence represents a S. pneumoniae ORF amino acid sequence from the
present invention. Note - The sequence data for this patent did not form
part of the printed specification, but was obtained in electronic format
directly from the USPTO web site.

Sequence 1529 AA;

Query Match 18.6%; Score 750.5; DB 9; Length 1529;
Best Local Similarity 79.7%; Pred. No. 1.6e-34;
Matches 149; Conservative 14; Mismatches 23; Indels 1; Gaps 1;

QY 1 KLGEIAESKFNKNGKESLKDITGVVHHQNEESIEKESFTIDRNFENK 60
Db 1336 KLGEIPESKFNKKNVDDSLNKETAENVNLLVDNQSIGKSLFNIHKTISTIRDFENK 1395

QY 61 DLKLLIKKKFREVDDFTSETGRMEEDYKYDDKGNIIAYDDGTDLLEYETEKLDIKSKI 120
Db 1396 DLKLLIKKKYKQEDDFVN-GGTRTVVERDYKYDDKGNIIAYDDGTDLLEYETEKLDIKSKI 1454

QY 121 YGVLSPSKDGHPFELIGKISNVSKNAKYVGNKYKIEIKATKYDFHSTMTFDLYANIND 180
Db 1455 YGVLSPSKDGHPFELIGKISNVSKNAKYVGNKYKIEIKATKYDFHSTMTFDLYANIND 1514

QY 181 IVDGLAF 187

Db 1515 IVDGISF 1521

RESULT 12

AAW55096

ID AAW55096 standard; protein; 117 AA.

XX AAW55096;

XX 02-OCT-1998 (first entry)

XX

DE Streptococcus pneumoniae SP043 protein.
XX
KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW detection; pneumonia; otitis media; meningitis.
XX
OS Streptococcus pneumoniae.
XX
PN W09818930-A2.
XX
XX 07-MAY-1998.
XX
XX 30-OCT-1997; 97WO-US019422.
XX
XX 31-OCT-1996; 96US-0029960P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Kunsch CA, Choi GH, Johnson LS, Hromockyj A;
PI
XX
DR WPI; 1998-272224/24.
DR N-PSDB; AAV27357.
XX
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae
PT - or their epitope-containing fragments, useful in protective or
PT therapeutic vaccines, and for diagnosis.
XX
XX Claim 11; Page 62; 118pp; English.
XX
XX The present sequence represents a protein from Streptococcus pneumoniae.
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC can be useful in vaccines for inducing protective antibodies against
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect Streptococcus infection (by usual hybridisation or
CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
CC (especially 10-300) mu g/ml per dose
XX
XX Sequence 117 AA;
SQ
Query Match 15.3%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.8e-28;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 657 YKGELEKGYQFDGWEISGFEGKKGAGYVNLKDTFIKPVFKIEEKKKEENKPTFDVSK 716
DB 1 YKGELEKGYQFDGWEISGFEGKKGAGYVNLKDTFIKPVFKIEEKKKEENKPTFDVSK 60
QY 717 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 773
DB 61 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 117
RESULT 13
ABP54590
ID ABP54590 standard; protein; 117 AA.
XX
AC ABP54590;
XX
DT 04-SEP-2002 (first entry)
XX
XX S. pneumoniae SP043 protein sequence SEQ ID NO:68.
DE
KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
KW antibacterial; Streptococcal infection; detection.
XX
OS Streptococcus pneumoniae.
XX

PN US2002061545-A1.
XX
PD 23-MAY-2002.
XX
XX 22-JAN-2001; 2001US-00765272.
PF
XX 30-OCT-1997; 97US-00961083.
PR
XX (CHOI/) CHOI G H.
PA (KUNS/) KUNSCH C A.
PA (BARA/) BARASH S C.
PA (DILL/) DILLON P J.
PA (DOUG/) DOUGHERTY B.
PA (FANN/) FANNON M R.
PA (ROSE/) ROSEN C A.
XX
PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
PI Rosen CA;
XX
XX WPI; 2002-479261/51.
DR N-PSDB; ABQ84825.
XX
XX New Streptococcus pneumoniae antigens, useful for detecting Streptococcus
PT and for preventing or attenuating disease caused by Streptococcus
PT infection.
XX
XX Claim 11; Page 29; 70pp; English.
XX
XX ABQ84792 to ABQ84904 represents nucleic acids which encode the
CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669. The S.
CC pneumoniae antigens have antibacterial activity and can be used in
CC vaccines. The S. pneumoniae antigens can also be used to prevent or
CC attenuate a Streptococcal infection in an animal. The polynucleotides
CC encoding the S. pneumoniae antigens can be used to detect Streptococcus
CC nucleic acids. ABQ84905 to ABQ85130 represent primers used in the cloning
CC of S. pneumoniae ORFs (open reading frames) which are used in an example
CC from the present invention
XX
XX Sequence 117 AA;
SQ
Query Match 15.3%; Score 615; DB 5; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.8e-28;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 657 YKGELEKGYQFDGWEISGFEGKKGAGYVNLKDTFIKPVFKIEEKKKEENKPTFDVSK 716
DB 1 YKGELEKGYQFDGWEISGFEGKKGAGYVNLKDTFIKPVFKIEEKKKEENKPTFDVSK 60
QY 717 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 773
DB 61 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTTNNPNK 117
RESULT 14
ADC45149
ID ADC45149 standard; protein; 117 AA.
XX
XX ADC45149;
AC
XX
DT 18-DEC-2003 (first entry)
XX
DE S. pneumoniae antigenic protein SPJ43.
XX
KW Antigen; bacterial infection; vaccine; pneumonia; antibacterial.
XX
OS Streptococcus pneumoniae.
XX
XX US6573082-B1.
PN
XX 03-JUN-2003.
PD
XX 28-MAR-2000; 2000US-00536784.
PP
XX

```
PR 31-OCT-1996; 96US-0029960P.
PR 30-OCT-1997; 97US-00961083.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
PI Rosen CA;
XX
XX WP1; 2003-764574/72.
DR N-PSDB; ADC45148.
XX
XX Novel polynucleotide encoding Streptococcus pneumoniae polypeptides
PT useful for producing vaccines for prevention or attenuation of infection
PT by Streptococcus pneumoniae.
XX
XX Example 1; SEQ ID NO 68; 58pp; English.
XX
XX The invention relates to an isolated polynucleotide consisting of a
CC Streptococcus pneumoniae nucleic acid (appearing as ADC45122 and encoding
CC SP028) one of 113 disclosed nucleic acids encoding 113 S. pneumoniae
CC antigens. Also included are making a recombinant vector by inserting the
CC nucleic acid into a vector, an isolated polynucleotide consisting of at
CC least 50 or 100 contiguous nucleotides of the SP028 nucleic acid, and a
CC recombinant host cell comprising the SP028 polynucleotide. The nucleic
CC acids are useful as DNA vaccine against Streptococcus pneumoniae
CC infection (e.g. pneumonia). Nucleic acids derived from the S. pneumoniae
CC antigen nucleic acids are useful as probes for use in diagnostic methods
CC for detecting S. pneumoniae gene expression. The present sequence
CC represents an S. pneumoniae antigenic protein.
XX
XX Sequence 117 AA;
SQ
Query Match 15.3%; Score 615; DB 7; Length 117;
Best Local Similarity 100.0%; Pred. No. 3.8e-28; Mismatches 0; Indels 0; Gaps 0;
Matches 117; Conservative 0;
QY 657 YKGELEKGYQFDGWEISGPEGKKGADGYVNLKSDTFIKPVFKKIEKKEENKPTFDVSK 716
Db 1 YKGELEKGYQFDGWEISGPEGKKGADGYVNLKSDTFIKPVFKKIEKKEENKPTFDVSK 60
QY 717 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 773
Db 61 KKDNPQVNHSQLNESHKEDLQREHSQKSDTKDVTATVLDKNNISSKSTNNPNK 117
RESULT 15
ABU24813
ID ABU24813 standard; protein; 1196 AA.
AC ABU24813;
DT 19-JUN-2003 (first entry)
DE Protein encoded by Prokaryotic essential gene #10340.
XX
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
XX Clostridium botulinum.
XX
XX WO200277183-A2.
PN
XX
XX 03-OCT-2002.
PD
XX
XX 21-MAR-2002; 2002WO-US009107.
PF
XX
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX
```

```
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI: 2003-029926/02.
DR N-PSDB; ACA28663.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 52737; 1766pp; English.
PS
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 1196 AA;
SQ
Query Match 5.6%; Score 227; DB 6; Length 1196;
Best Local Similarity 20.2%; Pred. No. 0.00025;
Matches 185; Conservative 145; Mismatches 260; Indels 328; Gaps 46;
QY 1 KLGEIAESKFKNLGNKGEG-----SLKKDVTGVHHHQBNEESI 39
Db 123 RLKDIQE-LFMDTGIGKEGYIIGGKIEAVLSGKPEERRSLLEAAGIVFKWKEAD 181
QY 40 KEKSFITDRNISTIRDFEN--KDLKKLIKKKFREVDFF--TSETGKME----- 85
Db 182 KKLNS--TEONLIRIKDILNTVEERMEPLKESEKAKKFLNLSELKKEKRVNMIYSIDK 239
QY 86 -BYDYKYDDKGNITAYDDGTD-LEYETEKLEDEIKSYIGVLSPKDGHFEILGKISNVSK 143
Db 240 TEKDLK-NISSNMLSLGENIDNLKNEKSQYKEIISKF-----NEKLELDK--NNSR 288
QY 144 NAKVYVGNYSIEIKATKYDFHSKTMFTFDLYANINDIVDGLAFAGDMRLFVKDNDQKA 203
Db 289 NKEEYVYVGNKDKDIE-----NENALKEKIKNLKD-----NIKVENTLKTN 331
QY 204 EIKIR-----MPEKIKETKSEYPVY----- 223
Db 332 EERLQKYLKEKEALEKKINKLKEESYLRBEIGKEDNVNKNELKEKEKILKSEE 391
QY 224 -----SSYGNVIELGEGDL--SKNKPDNUTKMGESKIYS-----DSEKQOYLK 265
Db 392 TEILSNTSLNRNEISVMENEIVNLKNDIKNSCDSYISSININIKTKDIEIKNIK 451
```



```
Qy 361 IDKGNLVTSSKLVLFKGDKEYTGDKFNVEAIKEDGSMFLFDTKPVNLSMDKNYFNP 420
Db 361 IDKGNLVTSSKLVLFKGDKEYTGDKFNVEAIKEDGSMFLFDTKPVNLSMDKNYFNP 420
Qy 421 SKSNKIYVRNPEFYLRGKISDKGFNWLNVESVDNYLYIGDLHIDNTRDFNKLNVK 480
Db 421 SKSNKIYVRNPEFYLRGKISDKGFNWLNVESVDNYLYIGDLHIDNTRDFNKLNVK 480
Qy 481 DGDIMDMGMDYKANGFPDKVTMDGNVYLQGYSDLNKAVGVHYQFLYDNVKEPVNID 540
Db 481 DGDIMDMGMDYKANGFPDKVTMDGNVYLQGYSDLNKAVGVHYQFLYDNVKEPVNID 540
Qy 541 PKGNTSIEYADGKSVFNFINDKRNNGFDGEIOEHIYINGKEYTSFNDIKQIIDKTLNIK 600
Db 541 PKGNTSIEYADGKSVFNFINDKRNNGFDGEIOEHIYINGKEYTSFNDIKQIIDKTLNIK 600
Qy 601 IVVKDFARNTTVKEFILNKDGEVSELKPHRVTVTIQNGKEMSTIYSEEDFILPVYKGE 660
Db 601 IVVKDFARNTTVKEFILNKDGEVSELKPHRVTVTIQNGKEMSTIYSEEDFILPVYKGE 660
Qy 661 LEKGYPQDGEWISGFEKGKADAGYVNLNSKDTFIKPVFKIIEKKEEENKPTFDVSKKDN 720
Db 661 LEKGYPQDGEWISGFEKGKADAGYVNLNSKDTFIKPVFKIIEKKEEENKPTFDVSKKDN 720
Qy 721 PQVNSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 773
Db 721 PQVNSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 773
```

RESULT 2

```
US-09-769-744A-28
; Sequence 28, Application US/09769744A
; Publication No. US20030134407A1
; GENERAL INFORMATION:
; APPLICANT: Le Page, Richard Wf
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hanniffy, Sean B
; APPLICANT: Hanebro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21122WO
; CURRENT APPLICATION NUMBER: US/09/769,744A
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: PCT/GB99/02452
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: GB 9816336.3
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125329
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 28
; LENGTH: 2119
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-744A-28
```

```
Query Match 100.0%; Score 4026; DB 3; Length 2119;
Best Local Similarity 100.0%; Pred. No. 7.6e-216;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KLGEIAESKFNVLNGKSGSLKKTGTGVEHHHQENESIKEKSSFTIDRNISTIRDPENK 60
Db 1313 KLGEIAESKFNVLNGKSGSLKKTGTGVEHHHQENESIKEKSSFTIDRNISTIRDPENK 1372
Qy 61 DLKXLIKKKFRVDDFTSETGRMEEDYKYDDKGNIIAYDDGTDLEVEKLEIKSKI 120
Db 1373 DLKXLIKKKFRVDDFTSETGRMEEDYKYDDKGNIIAYDDGTDLEVEKLEIKSKI 1432
Qy 121 YGVLSPSKDGHEILGKISNVSQNAKYVYGNYSIEIKATKYDFHSKTMFTFDLYANIND 180
Db 1433 YGVLSPSKDGHEILGKISNVSQNAKYVYGNYSIEIKATKYDFHSKTMFTFDLYANIND 1492
```

```
Qy 181 IVDGLAPAGDMRLFVKNDOKKAEIKIRMEPEKIKETKSEYPYVSSYGNVIELGEGDLSKN 240
Db 1493 IVDGLAPAGDMRLFVKNDOKKAEIKIRMEPEKIKETKSEYPYVSSYGNVIELGEGDLSKN 1552
Qy 241 KPNLNTWBSGKLYSDSEKQOYLLKNNIILRGYALAVTTYNPGKTDMLGNGVYSKEDI 300
Db 1553 KPNLNTWBSGKLYSDSEKQOYLLKNNIILRGYALAVTTYNPGKTDMLGNGVYSKEDI 1612
Qy 301 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFKNDKGEA 360
Db 1613 AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFKNDKGEA 1672
Qy 361 IDKGNLVTSSKLVLFKGDKEYTGDKFNVEAIKEDGSMFLFDTKPVNLSMDKNYFNP 420
Db 1673 IDKGNLVTSSKLVLFKGDKEYTGDKFNVEAIKEDGSMFLFDTKPVNLSMDKNYFNP 1732
Qy 421 SKSNKIYVRNPEFYLRGKISDKGFNWLNVESVDNYLYIGDLHIDNTRDFNKLNVK 480
Db 1733 SKSNKIYVRNPEFYLRGKISDKGFNWLNVESVDNYLYIGDLHIDNTRDFNKLNVK 1792
Qy 481 DGDIMDMGMDYKANGFPDKVTMDGNVYLQGYSDLNKAVGVHYQFLYDNVKEPVNID 540
Db 1793 DGDIMDMGMDYKANGFPDKVTMDGNVYLQGYSDLNKAVGVHYQFLYDNVKEPVNID 1852
Qy 541 PKGNTSIEYADGKSVFNFINDKRNNGFDGEIOEHIYINGKEYTSFNDIKQIIDKTLNIK 600
Db 1853 PKGNTSIEYADGKSVFNFINDKRNNGFDGEIOEHIYINGKEYTSFNDIKQIIDKTLNIK 1912
Qy 601 IVVKDFARNTTVKEFILNKDGEVSELKPHRVTVTIQNGKEMSTIYSEEDFILPVYKGE 660
Db 1913 IVVKDFARNTTVKEFILNKDGEVSELKPHRVTVTIQNGKEMSTIYSEEDFILPVYKGE 1972
Qy 661 LEKGYPQDGEWISGFEKGKADAGYVNLNSKDTFIKPVFKIIEKKEEENKPTFDVSKKDN 720
Db 1973 LEKGYPQDGEWISGFEKGKADAGYVNLNSKDTFIKPVFKIIEKKEEENKPTFDVSKKDN 2032
Qy 721 PQVNSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 773
Db 2033 PQVNSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 2085
```

RESULT 3

```
US-10-282-122A-73670
; Sequence 73670, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Havelbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
```

```
/
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 73670
; LENGTH: 2140
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-282-122A-73670

Query Match          100.0%; Score 4026; DB 4; Length 2140;
Best Local Similarity 100.0%; Pred. No. 7.7e-216;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGEIAESKFNGLNGKEGSLKDDTTGVVHHQNEESIEKSSFTIDRNISTIRDFENK 60
Db 1334 KLGEIAESKFNGLNGKEGSLKDDTTGVVHHQNEESIEKSSFTIDRNISTIRDFENK 1393

QY 61 DLKLLIKKKFREVDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLLEYETEKLEIKSKI 120
Db 1394 DLKLLIKKKFREVDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLLEYETEKLEIKSKI 1453

QY 121 YGVLSPSKDGHEFELGKISNVSKNAKVYGNKYKSIIEIKATKYDPSHKTMTFDLYANIND 180
Db 1454 YGVLSPSKDGHEFELGKISNVSKNAKVYGNKYKSIIEIKATKYDPSHKTMTFDLYANIND 1513

QY 181 IVDGLAFAGDMRLFKVNDQKAEIKIRMEPKIKETKSEYPVYSSYGVNIELEGDLSKN 240
Db 1514 IVDGLAFAGDMRLFKVNDQKAEIKIRMEPKIKETKSEYPVYSSYGVNIELEGDLSKN 1573

QY 241 KPDNLTKMESGKIYSDSEKQQYLLKDNIIILRGYALKVTTYNPGKTDMLGNGVYSKEDI 300
Db 1574 KPDNLTKMESGKIYSDSEKQQYLLKDNIIILRGYALKVTTYNPGKTDMLGNGVYSKEDI 1633

QY 301 AKIQANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGNIIIRYQVFTFQNDKGEA 360
Db 1634 AKIQANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGNIIIRYQVFTFQNDKGEA 1693

QY 361 IDKGNLVTDSSKLVLFCKDKKEYTGEDKFNVEAIEKDGSMFLFIDTKPVNLSMDKNYFNP 420
Db 1694 IDKGNLVTDSSKLVLFCKDKKEYTGEDKFNVEAIEKDGSMFLFIDTKPVNLSMDKNYFNP 1753

QY 421 SKSNKIYVRNPFYLRGKISDKGGFNWELRVNESVVDNYLIYGDLDHINTDRDNFKLVNK 480
Db 1754 SKSNKIYVRNPFYLRGKISDKGGFNWELRVNESVVDNYLIYGDLDHINTDRDNFKLVNK 1813

QY 481 DGDIMDWGMKDYKANGFPDVKVTDMDGNVYLOTGYSDLNAAKVGVHYQFLYDNVKEPVNID 540
Db 1814 DGDIMDWGMKDYKANGFPDVKVTDMDGNVYLOTGYSDLNAAKVGVHYQFLYDNVKEPVNID 1873

QY 541 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIOEQHIYINGKEYTSFNDIKQIIDKTLNIK 600
Db 1874 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIOEQHIYINGKEYTSFNDIKQIIDKTLNIK 1933

QY 601 IVVKDFARNTTVKFEILNKDTGEVSELKPHRVTVVTIQNGKEMSSTIVSEEDFILPVYKGE 660
Db 1934 IVVKDFARNTTVKFEILNKDTGEVSELKPHRVTVVTIQNGKEMSSTIVSEEDFILPVYKGE 1993

QY 661 LEKGYQFDGWEISGPEGKADAGVINLSKDTFKPVFKKIEPKKEENKPTFDVSKKXDN 720
Db 1994 LEKGYQFDGWEISGPEGKADAGVINLSKDTFKPVFKKIEPKKEENKPTFDVSKKXDN 2053

QY 721 PQVNHSQLNESHKEDLQREEHSQKSDSTKDVATVLDKNNISSKSTTNNPK 773
Db 2054 PQVNHSQLNESHKEDLQREEHSQKSDSTKDVATVLDKNNISSKSTTNNPK 2106
```

RESULT 4

```
US-10-472-928-1180
; Sequence 1180, Application US/10472928
; Publication NO. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926W0
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 1180
; LENGTH: 2140
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: serine protease, subtilase family
; OTHER INFORMATION: Cellular location: Peptidoglycan-bound (LPXTG)
; OTHER INFORMATION: Similar to strain R6 sequence 15902605 (0.E+01)
US-10-472-928-1180

Query Match          100.0%; Score 4026; DB 5; Length 2140;
Best Local Similarity 100.0%; Pred. No. 7.7e-216;
Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KLGEIAESKFNGLNGKEGSLKDDTTGVVHHQNEESIEKSSFTIDRNISTIRDFENK 60
Db 1334 KLGEIAESKFNGLNGKEGSLKDDTTGVVHHQNEESIEKSSFTIDRNISTIRDFENK 1393

QY 61 DLKLLIKKKFREVDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLLEYETEKLEIKSKI 120
Db 1394 DLKLLIKKKFREVDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLLEYETEKLEIKSKI 1453

QY 121 YGVLSPSKDGHEFELGKISNVSKNAKVYGNKYKSIIEIKATKYDPSHKTMTFDLYANIND 180
Db 1454 YGVLSPSKDGHEFELGKISNVSKNAKVYGNKYKSIIEIKATKYDPSHKTMTFDLYANIND 1513

QY 181 IVDGLAFAGDMRLFKVNDQKAEIKIRMEPKIKETKSEYPVYSSYGVNIELEGDLSKN 240
Db 1514 IVDGLAFAGDMRLFKVNDQKAEIKIRMEPKIKETKSEYPVYSSYGVNIELEGDLSKN 1573

QY 241 KPDNLTKMESGKIYSDSEKQQYLLKDNIIILRGYALKVTTYNPGKTDMLGNGVYSKEDI 300
Db 1574 KPDNLTKMESGKIYSDSEKQQYLLKDNIIILRGYALKVTTYNPGKTDMLGNGVYSKEDI 1633

QY 301 AKIQANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGNIIIRYQVFTFQNDKGEA 360
Db 1634 AKIQANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGNIIIRYQVFTFQNDKGEA 1693

QY 361 IDKGNLVTDSSKLVLFCKDKKEYTGEDKFNVEAIEKDGSMFLFIDTKPVNLSMDKNYFNP 420
Db 1694 IDKGNLVTDSSKLVLFCKDKKEYTGEDKFNVEAIEKDGSMFLFIDTKPVNLSMDKNYFNP 1753

QY 421 SKSNKIYVRNPFYLRGKISDKGGFNWELRVNESVVDNYLIYGDLDHINTDRDNFKLVNK 480
Db 1754 SKSNKIYVRNPFYLRGKISDKGGFNWELRVNESVVDNYLIYGDLDHINTDRDNFKLVNK 1813

QY 481 DGDIMDWGMKDYKANGFPDVKVTDMDGNVYLOTGYSDLNAAKVGVHYQFLYDNVKEPVNID 540
Db 1814 DGDIMDWGMKDYKANGFPDVKVTDMDGNVYLOTGYSDLNAAKVGVHYQFLYDNVKEPVNID 1873

QY 541 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIOEQHIYINGKEYTSFNDIKQIIDKTLNIK 600
Db 1874 PKGNTSIEYADGKSVVFNINDKRNNGFDGEIOEQHIYINGKEYTSFNDIKQIIDKTLNIK 1933

QY 601 IVVKDFARNTTVKFEILNKDTGEVSELKPHRVTVVTIQNGKEMSSTIVSEEDFILPVYKGE 660
```


Db 1934 IVKDFARNTTVKGFILNKDTGEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKGE 1993
Qy 661 LEKYQDFGWEISGEGKADAGYVINLSKDTFIKPVFKKIEEKEEENKPTFDVSKKKDN 720
Db 1994 LEKYQDFGWEISGEGKADAGYVINLSKDTFIKPVFKKIEEKEEENKPTFDVSKKKDN 2053
Qy 721 PAVNHSQLNESHKREDLQREHSHQSDSTKDVDTATVLDKNNISSKSTTNNPNK 773
Db 2054 PAVNHSQLNESHKREDLQREHSHQSDSTKDVDTATVLDKNNISSKSTTNNPNK 2106
RESULT 5
US-10-617-320-3169
; Sequence 3169, Application US/10617320
; Publication No. US20050136404A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/617,320
; FILING DATE: 10-Jul-2003
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3169:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...637
; SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
US-10-617-320-3169
Query Match 75.7%; Score 3048; DB 5; Length 637;
Best Local Similarity 99.3%; Pred. No. 6.5e-162;
Matches 585; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Qy 185 LAFAGDMRLFVKDNDQKAEIKIRMPKIKETKSEYPVSSYGNVIELGEGDLSKNKPDN 244
Db 15 LAFAGDMRLFVKDNDQKAEIKIRMPKIKETKSEYPVSSYGNVIELGEGDLSKNKPDN 74

Qy 245 LTKWESKIIYSDSEKQOYLLKDNILIRKGYALKVTTYNPGKTDMLGNGVYSKEDIKIQ 304
Db 75 LTKWESKIIYSDSEKQOYLLKDNILIRKGYALKVTTYNPGKTDMLGNGVYSKEDIKIQ 134
Qy 305 KANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFKMNDKGEAIDKD 364
Db 135 KANPNLRALSETTIYADSRNVEDGRSTQAVLMSALDGFNIIRYQVFTFKMNDKGEAIDKD 194
Qy 365 GNLVTSSKLVLPFGKDDKEYTGEDKFNVEAIKEDGSMFLFDTTKPVNLSMDKNVPNPKSN 424
Db 195 GNLVTSSKLVLPFGKDDKEYTGEDKFNVEAIKEDGSMFLFDTTKPVNLSMDKNVPNPKSN 254
Qy 425 KIYVRNPEFYLRGKISDKGGFNWELFVNESVVDNYLIYGLDHIHNDTRDFNIKLVKDGDI 484
Db 255 KIYVRNPEFYLRGKISDKGGFNWELFVNESVVDNYLIYGLDHIHNDTRDFNIKLVKDGDI 314
Qy 485 MDGMKDYKANGFPDKVTDMDGNVYLQTYGSDLNAKAVGVHYQFLYDYNVKEPVNIDPKGN 544
Db 315 MDGMKDYKANGFPDKVTDMDGNVYLQTYGSDLNAKAVGVHYQFLYDYNVKEPVNIDPKGN 374
Qy 545 TSIEYADGKSVVFNINDKRNNGFDGEIQEOHIIYNGEYTSFNDIKOIIDKTLNIIKIVVK 604
Db 375 TSIEYADGKSVVFNINDKRNNGFDGEIQEOHIIYNGEYTSFNDIKOIIDKTLNIIKIVVK 434
Qy 605 DFARNTTVKEFIILNKDTGEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKGELEKG 664
Db 435 DFARNTTVKEFIILNKDTGEVSELKPHRVTVTIQNGKEMSSSTIVSEEDFILPVYKGELEKG 494
Qy 665 YQFDGWEISGFEKGKADAGYVINLSKDTFIKPVFKKIEEKEEENKPTFDVSKKKDNQVNV 724
Db 495 YQFDGWEISGFEKGKADAGYVINLSKDTFIKPVFKKIEEKEEENKPTFDVSKKKDNQVNV 554
Qy 725 HSQLNESHKREDLQREHSHQSDSTKDVDTATVLDKNNISSKSTTNNPNK 773
Db 555 HSQLNESHKREDLQREHSHQSDSTKDVDTATVLDKNNISSKSTTNNPNK 603
RESULT 6
US-10-617-320-4771
; Sequence 4771, Application US/10617320
; Publication No. US20050136404A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/617,320
; FILING DATE: 10-Jul-2003
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3169:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...637
; SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
US-10-617-320-3169
Query Match 75.7%; Score 3048; DB 5; Length 637;
Best Local Similarity 99.3%; Pred. No. 6.5e-162;
Matches 585; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
Qy 185 LAFAGDMRLFVKDNDQKAEIKIRMPKIKETKSEYPVSSYGNVIELGEGDLSKNKPDN 244
Db 15 LAFAGDMRLFVKDNDQKAEIKIRMPKIKETKSEYPVSSYGNVIELGEGDLSKNKPDN 74

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4771:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1529 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1529
; SEQUENCE DESCRIPTION: SEQ ID NO: 4771:
US-10-617-320-4771

Query Match 18.6%; Score 750.5; DB 5; Length 1529;
Best Local Similarity 79.7%; Pred. No. 3.6e-33;
Matches 149; Conservative 14; Mismatches 23; Indels 1; Gaps 1;

QY 1 KLGEIAESFKNLGKGLKDTGTVEHHQENEESIKESSTPTIDRNISTIRDFENK 60
Db 1336 KLGEIPESFKNLKKNVKDSLNRKETAENVNLLVONQSLFNHKTISTIRDFENK 1395

QY 61 DLKLLIKKKFREVDFTSETGKRMEEYDYKDDKGNIIAYDDGTDLLEYETEKLDIKSKI 120
Db 1396 DLKLLIKKKYKQEDFVN-GGRTVERDYKDDKGNIIAYDDGTDLLEYETEKLDIKSKI 1454

QY 121 YGVLSPKDGHFEILGKISNVSKNAKYVYGNKYKSIETKATKYDFHSKTMTFPLVANIND 180
Db 1455 YGVLSPKDGHFEILGKISNVSKNAKYVYGNKYKSIETKATKYDFHSKTMTFPLVANIND 1514

QY 181 IVDGLAF 187
Db 1515 IVDGISF 1521

RESULT 7
US-09-765-272-68
; Sequence 68, Application US/09765272
; Patent No. US20020061549A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512

```

```

; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-272-68

Query Match 15.3%; Score 615; DB 3; Length 117;
Best Local Similarity 100.0%; Pred. No. 5e-27;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YKGELEKGYQFDGWEISGFEGKKGADAGYVINLSKDTFIKPVFKKIEEKEENKPTFDVSK 716
Db 1 YKGELEKGYQFDGWEISGFEGKKGADAGYVINLSKDTFIKPVFKKIEEKEENKPTFDVSK 60

QY 717 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPK 773
Db 61 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPK 117

RESULT 8
US-11-106-649-68
; Sequence 68, Application US/11106649
; Publication No. US20050181439A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; FILE REFERENCE: PB340P2C3D1
; CURRENT APPLICATION NUMBER: US/11/106,649
; CURRENT FILING DATE: 2005-04-15
; PRIOR APPLICATION NUMBER: US 09/765,271
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: US 09/536,784
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: US 08/961,083
; PRIOR FILING DATE: 1997-10-30
; PRIOR APPLICATION NUMBER: US 60/029,960
; PRIOR FILING DATE: 1996-10-31
; NUMBER OF SEQ ID NOS: 454
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 68
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-11-106-649-68

Query Match 15.3%; Score 615; DB 6; Length 117;
Best Local Similarity 100.0%; Pred. No. 5e-27;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YKGELEKGYQFDGWEISGFEGKKGADAGYVINLSKDTFIKPVFKKIEEKEENKPTFDVSK 716
Db 1 YKGELEKGYQFDGWEISGFEGKKGADAGYVINLSKDTFIKPVFKKIEEKEENKPTFDVSK 60

QY 717 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPK 773
Db 61 KKDNPQVNHSQLNESHKEDLQREHSQKSDSTKDVATATVLDKNNISSKSTTNNPK 117

RESULT 9
US-10-732-923-3304
; Sequence 3304, Application US/10732923
; Publication No. US20050108791A1
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 38-15(52796)C
; CURRENT APPLICATION NUMBER: US/10/732,923
; CURRENT FILING DATE: 2003-12-10
; PRIOR APPLICATION NUMBER: 10/310,154

```



```
QY 86 -EYDYKDDKGNIIAYDDGTD-LEYETEKLEDEIKSKIYGLSPSKDGHFEILGKISNVSK 143
Db 240 IEKDLK-NISSNLSIGENIDNLKNEKSQYKEIISKF-----NEKLELLDK--NNSR 288
QY 144 NAKVYGNKYKSEIKATKYDPHSHKTTWTFDLVANINDIVDGLAFAGDMRLFVKNDQOKA 203
Db 289 NKEEYNNKDKDKDIE-----NENALLKEKIKNLKD-----NIKVKENTLKTN 331
QY 204 EIKIR-----MPEKIKETSEYPV-----223
Db 332 EERLOKYLKEALEKINKLKEESYLREIECKEDNVNPNFKELKEKEKLIKLSKE 391
QY 224 -----SYGNVIELGEGDL--SKVKPNLTKMESGKIYS-----DSEKQOYLK 265
Db 392 EIBLSNTSLRNEISWENEIVLENKLDNIKNKSCDSYISSININIKTKEDIEKEIKNIK 451
QY 266 DNILL-----RKGYALKVTTYNPCKTMDLEGNGVYSKED-----IAKTQK----AN 307
Db 452 ENILLLENNLKNRSNMISSLKISLNNKEK-KLKEKNAAYSRLLEANVHMLSNLEKHVEGYN 510
QY 308 PNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVTFPMNDKGBAIDKGNL 367
Db 511 RSVKTLMEHVSQKVDNIKGG-----CEVLGDIKVKKELETAMEIALGGAI---SNV 560
QY 368 VTDSSKLVLFGKDDKEYTGEDKFNVEAIKEDGSMFLFIDTKPVNLSMDKNYFNP---SKSN 424
Db 561 ITEDE-----NKAKILINLYLKKK-----SLGRATFLPLTTIQOR 594
QY 425 KIYVRNPEFLRGKISDKGPFWEELRVNESVVDNLYIYCDLHIDNTRDNI---KLNVDKG 482
Db 595 KAKINN-----VTREDGF---LGIASDLIDY-----DVKFSNIIYVLGRTLVAKD- 637
QY 483 DTMWGMKDKYKANGPDKVTMDGNVYLQTVGYSDLNKAKAVGVHYQFLYDNVKNPEVNIDPK 542
Db 638 ---MDSALKIATKLNYSFKVLTEGEV-----INPG 665
QY 543 GNT-----SIEYADGKSVFNINDKRNNGFDGEIOEQHIYINGKEYTSFNDIKIIDKTANI 599
Db 666 GSLTGSIKHRAGSSI---ISKR-----EIEE-----TKKELEBETKN- 700
QY 600 KIVVKDFARNTTVKFEILNKDTGVESELKPHRVTVTIQNGKEMSTIYSEEDFILPVYKG 659
Db 701 -----FIEBFMGN-----ILENKNKIKT--LDEENLNI---KD 728
QY 660 ELEKGYQFGWEISGFEGKDA-----GYVINLSKD--TFIKPVFKIIEKKEEENK 709
Db 729 EI-----YNNIEITKPMGKLNAIKEDTERLRSLSLISREIEIKLTOKDKIDEEENINESQK 784
QY 710 PTFDVSKKKDNQVNHSQLSHRKEDELDQREHSQKSDTK-----DVTATVLDKNNISSK 765
Db 785 QLEELKLKDD---LNHNDIKEC--EDFLQNEBENVKNIKDKLIEYKIEKAKLDLMLVSIK 839
QY 766 -----STNNPNK 773
Db 840 KELYSMDTNIITLNNENK 857

RESULT 11
US-09-820-843A-109
; Sequence 109, Application US/09820843A
; Publication No. US2003039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 109
; LENGTH: 861
```

```
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: hypothetical protein
; NAME/KEY: misc_feature
; OTHER INFORMATION: gi|4493994
US-09-820-843A-109

Query Match          5.5%; Score 222; DB 3; Length 861;
Best Local Similarity 19.3%; Pred. No. 0.00055;
Matches 162; Conservative 136; Mismatches 324; Indels 216; Gaps 37;

QY 2 LGEIAESFKNLGNG-----KEGSLKDDTTGVEHHH--QENEESIEKSSFTI 47
Db 154 LNNINGKFKNLDDCYNLPTNNLYIDKEG--KMLTKEHYNAASSNEYNNKNNY 211
QY 48 DRNISTIRDFENKOL-----KKLIKPKPREVDFTSETGRMEEEYDYKDDKGNIT 98
Db 212 NNNSYNNNFNCNNYNDNNYNNNNKMGKYSRLNVLK--KEHDMVDYEYNNKGNIR 268
QY 99 AYD-----DGTDEYETEKLDEIKSKIYCVLSPSKDGHPEI--LGKISNVSKNAKYVYN 151
Db 269 KNDSEKYWDNPLHY-----SKQNNYDIFTLGDIKKYAKNKECKGN 310
QY 152 NYKSEIKATKYDPHSHKTTWTFDLVANINDIVDG---LAFAGDMRLFVKNDQOKAEIKIR 208
Db 311 KYMMN-----HDNNSN-----NSNNVLNNNNNSNNYNNIFKNDDENL-----351
QY 209 MPEKIKETKSEYPVSYGNVIELGEG-DLSK---NKPDLNLTMBESGKIYSDSEKQOYL 264
Db 352 -----TKSNFAKWFKNNNNNVNVNTDIIKLANK-----NSQGHSDGKNNNNN 396
QY 265 KDNILRGYALKVTTYNPCKTMDLEGNGVYSKEDIKIQKANPN---LRALSETTIYAD 321
Db 397 GNNII-----NNNSNNKNIIFQGNRSRYENVYNNIINNNNNNIISNNKNEASFND 447
QY 322 SRNVEDGRSTQSVLMSALDGFNIIRYQVTFPMNDKG--EADKDGNLVTDSSKLVLFGK 380
Db 448 NINTNSGREEEKISNTVAE---LLMKQISMIKERNKGLDVLKKT-------FGFL 494
QY 381 DKEYTGEDKFNVEAIKEDGSMFLFIDTKPVNLSMDKNYFNPSKSKIYVRNPEFL- 435
Db 495 DNNYQNYGSNN-----NSSLEKNNM---KENDIYSKEASKRIMDI 533
QY 436 ---RGKISDKGFWELRVNESVVDNLYIYCDLHIDNTRDNIKLVKDGDMWGMKD 491
Db 534 TLNNGSLVSQE-----SLLVNQSVLNNNNNNYNNNNNNNNNNNNNNNNNNNN 582
QY 492 YKANGFPDKVTDMDGNVYLQTVGYSDLNKAK---AVGVHYQFLYDNVKNPEVNDPKGNTSIE 548
Db 593 -----NNMNSNNNINNINNYYKNNHKYHSHMDNVTYKKIP--INNYSNNDGNNNSNS 633
QY 549 YADGKSVFNINDKRNNGFDGEIOEQHIYINGKEYTSFNDIKQIIDK-----595
Db 634 NSNNVHEYNNKKN--FKNKINNYHNLDPKNKNNMNNNTYNNINKNNLSNMENFPFSL 691
QY 596 TLNLIKIVVKDFARN---TTVKEFTLANKDTGEVSELKPHRVTVTIQNGKEMSTIYSEED 651
Db 692 SPFNSDINKNNAQGNINITPIINSILRLD-NEVDNVHNNISSENIONAK--VSNVLDLSK 748
QY 652 FILPVYKGELEKGYQFGWEISGFEGKDKDAGVNLNSKDTFIK-----PVFKIEEEK 704
Db 749 SLLKAKSKQGNNTNIP-----KNFNNNN-----NNNSKFINYNSQQYFSHQOQOQH 799
QY 705 EENKPTFDVSKKDNQVNHSQLSHRKEDELDQREHSQK---SDSTKDVATVLDK 759
Db 800 QOQOQOQ--QOQOTLIQTQINGTHLNDFNKKFKKERYPMKYPEFDGTTNETMVMVREK 855

RESULT 12
US-10-732-923-3352
; Sequence 3352, Application US/10732923
```


QY 481 -----DGDIMDGMKDYK-----ANGFPDKV 501
Db 691 TQATTKPGQAAGSALGDSVQAQAEQQAQPPVPPVPEAKAQVPTPPAPVNNKTENV 750
QY 502 TMDGNVYLOTGYSDLA-----KAVGVHYQFLYDNVKNPEVNIIDPKNGTSIEYADGKSVV 556
Db 751 SKLD---YLEKLYEFLNTSYICHKYLVSHTMNEKILKQYKTKTEESKLSLSCDPLDLL 807
QY 557 FNIN-----DKRNGFDG---EIOEHIYINGKEYTSFNDIKQIIDTKLAIKIVV 603
Db 808 FNIQNNIPVMYSMFDSLNSLSOLFMEIYKEMVNCNLYKLNKNDKILNLEBAKKVTSV 867
QY 604 KDFARNTTVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGELEK 663
Db 868 KTLSSSSMQPLSLTPQDKPEVSANDDTSHSTNLNLSLKLFPENILS-----LGKKN 918
QY 664 GYQFDGWEISGFEK-----KQAGYVINLSKDTFIKPVFKKTEEEKKEENKPTF--D 713
Db 919 IYQ---ELIGQSSNFYKILKDSDTFYNESFTNFVKSADDTINSLNDESKRKKLEED 974
QY 714 VSKKNDPNVHNSQLNESHREKDLQREHHSQKSDSTKDVATVLDKNNISSK-STNNP 771
Db 975 INKLKKTQLSFDLYNKKYKLERLFDKKKTGVKYMQIKKLTLLKEQLESKLNSLNNP 1033
RESULT 14
US-10-954-924-5
; Sequence 5, Application US/10954924
; Publication No. US20050095256A1
; GENERAL INFORMATION:
; APPLICANT: Bujard, Hermann
; APPLICANT: TOLLE, RALF
; APPLICANT: PAN, WEIQING
; TITLE OF INVENTION: Recombinant Process for Preparing a Complete Malaria Antigen GP19
; CURRENT APPLICATION NUMBER: GRUE-003DIV
; CURRENT FILING DATE: 2004-09-29
; PRIOR APPLICATION NUMBER: 09/269,874
; PRIOR FILING DATE: 1999-08-02
; PRIOR APPLICATION NUMBER: PCT/EP97/05441
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: DE 19640817.2
; PRIOR FILING DATE: 1996-10-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1621
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-954-924-5
Query Match 5.4%; Score 218; DB 5; Length 1621;
Best Local Similarity 18.6%; Pred. No. 0.002;
Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;
QY 23 KDTTG-VEHHQNEESIK-----EKSSFTIDRNISTRDPENKDLKLIKKEPREVDD 75
Db 232 KONGVMEDIYIKNNKKTIIENINELIEESKKTIDKKNATKEE-----KKKLYQAY 283
QY 76 FTSETGKRMEEYDYDDKGNIIAYDDGDTLLEYETEKLEDEIKSIYGVLSPSKOGHFEIL 135
Db 284 DLSIYNKQLEE-----AHNLISV-----LE---KRIDTLK-----KQENIKELL 319
QY 136 GKISNVSKNAKYVYGNKYKSIIEIKATKYDFH-----SKWTFPDYANINDIVDGLAPA 188
Db 320 DKINEIKNPPANGSNTWNTLLDKNKKEHEKEIKETAKTIKFNIDSLFTDPL----- 373
QY 189 GDMRLFVKDNDOKKAEIKRMPEKIKETKSEYP-----YVSSGVN-----IELGE-GDLSK 239
Db 374 -ELEYLDREKN-KNIDISAKVETKSEFPNPGVTPPLSLNINNALNELNSFGDLI- 430
QY 240 NKPDNLTKWESKSIYSDSEKQYL--LKDNIIL-----RKGVALKVTTYNPGKTDML 289

Db 431 -NPPDYTKEPSKNIYTDNERKKFINEIKKIKIEKKKTESDKSKSLN-----DIT 485
QY 290 EG-----NGVYSKE-----DIAKTQKA-----NPNLRALSETTIYADSRNVEDGRTQSV 334
Db 486 KEYEKLALNEIYDSKFNNIDLTNFEKMMGKRYSYKVEKLTHTHTTFSYENSXHNLEKLT 545
QY 335 LMSALDGF---NII-----RYQVFTFKMNDKGEAIDKQGNLVTDSKLVLFQKDDKEVT 385
Db 546 ALKYMEDYSLNIVVEKELKYKKNLSKIEHIEITLVE--NIKKOEEQLF-----EKKIT 598
QY 386 GEDFNVBAIKEDGSMFLFIDTKPVNLSMDKNYFNPSKSNKIYVRNPEFYLRGKISDKGKF 445
Db 599 KQENKPEDEKILEVSDIVKVQVQKVL--MKNIDELKKTQLILKNVE--LKHNIHVPNSY 653
QY 446 NWELRVNESVDNLYIYGLDHNIDTRDFNIK-----NVK----- 480
Db 654 KOENKQE---PYLLIVLKKEIDKLKVFMPKVESLINEEKKNIKTGEGSDNSEPTEGEI 709
QY 481 -----DGDIMDGMKDYK-----ANGFPDKV 501
Db 710 TQATTKPGQAAGSALGDSVQAQAEQQAQPPVPPVPEAKAQVPTPPAPVNNKTENV 769
QY 502 TMDGNVYLOTGYSDLA-----KAVGVHYQFLYDNVKNPEVNIIDPKNGTSIEYADGKSVV 556
Db 770 SKLD---YLEKLYEFLNTSYICHKYLVSHTMNEKILKQYKTKTEESKLSLSCDPLDLL 826
QY 557 FNIN-----DKRNGFDG---EIOEHIYINGKEYTSFNDIKQIIDTKLAIKIVV 603
Db 827 FNIQNNIPVMYSMFDSLNSLSOLFMEIYKEMVNCNLYKLNKNDKIKNLEBAKKVTSV 886
QY 604 KDFARNTTVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSSITVSEEDFILPVYKGELEK 663
Db 887 KTLSSSSMQPLSLTPQDKPEVSANDDTSHSTNLNLSLKLFPENILS-----LGKKN 937
QY 664 GYQFDGWEISGFEK-----KDAGYVINLSKDTFIKPVFKKTEEEKKEENKPTF--D 713
Db 938 IYQ---ELIGQSSNFYKILKDSDTFYNESFTNFVKSADDTINSLNDESKRKKLEED 993
QY 714 VSKKNDPNVHNSQLNESHREKDLQREHHSQKSDSTKDVATVLDKNNISSK-STNNP 771
Db 994 INKLKKTQLSFDLYNKKYKLERLFDKKKTGVKYMQIKKLTLLKEQLESKLNSLNNP 1052
RESULT 15
US-10-087-464-10
; Sequence 10, Application US/10087464
; Publication No. US20030059436A1
; GENERAL INFORMATION:
; APPLICANT: Chishti, Athar
; APPLICANT: Oh, Steven
; APPLICANT: Liu, David
; APPLICANT: Goel, Vikas
; APPLICANT: Li, Xuerong
; TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
; CURRENT APPLICATION NUMBER: US/10/087,464
; CURRENT FILING DATE: 2002-03-01
; PRIOR FILING DATE: 2002-03-01
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 1639
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-10-087-464-10
Query Match 5.4%; Score 218; DB 4; Length 1639;
Best Local Similarity 18.6%; Pred. No. 0.0021;
Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;
QY 23 KDTTG-VEHHQNEESIK-----EKSSFTIDRNISTRDPENKDLKLIKKEPREVDD 75

Db	232	KONVGMEDYIKNKKTJENINELIEESKKTIDKNKNAKEE-----KKGYQAOY 283
Qy	76	FTSETKRMEEYDYDDKGNIIAVDDGTDLEYETEKLDBIKSKIYGVLSPSKDGHFEL 135
Db	284	DIUSYKQLEF-----AHNLISV-----LE---KKIDTLK-----KXENIKEL 319
Qy	136	GKISNVSKNAKYVYDGNYSKISIEIKATKYDPH-----SKTMTFDYANINDIVDGLAFA 188
Db	320	DKXINEIKPPANGSNTPTLLDKNKIEEKEIKBIAKTIKFNIDSLFTDPL----- 373
Qy	189	GDMLFVKDNDOKKAEIKIRPEKIKETKSYP-----YVSSYGNV-----IELGE-CDLSK 239
Db	374	-ELEYLRKN-KNIDISAKVETKESTPEYNGVTPYPLSYDINDINNALMELNSFGDLI- 430
Qy	240	NKPDNLTKMESKIIYDSKEKOYL--LKDNIL-----RKGYALKVTTYNPGKTDML 289
Db	431	-NPFDTKEPSKNIYDNERKKFNEIKEKIEKKIESDKSYEDRSKSLN-----DIT 485
Qy	290	EG-----NGVYSKE-----DIAKTQKA-----NPNLRALSETTIYADSRNVEDGRSTQSV 334
Db	486	KEYEKLNEIYDSKFNNNIDLTNPEKMMGKRSYKVEKLTHTNTFASYENSKHLEKLT 545
Qy	335	LMSALDGP---NII-----RYQVPTFMNDKGEAIDKGNLYTDSKLVLPCKDDKEVT 385
Db	546	ALKYMEDYSLRNIYVEKELKYFNKLSKIENIEITLVE--NKKDBEQLF-----EKKT 598
Qy	386	GEDKFNVEAIKBDGSMFLDTDKPNLSMDKNYPNPSKSNKIYVRNPFYLRGKISDKGGF 445
Db	599	KDENKDEKILEVSDIVKVOQKVL---MKNIDELKKTQILKQVE--LKXNIHVPSY 653
Qy	446	NWELRVNESVDNYLIYDGLHDIDNTRDNKL-----NVK----- 480
Db	654	KQENKQE---PYLIVLKEIDKLKVPMPKVESLINEBKKNIKTEGSDNSEPSTEI 709
Qy	481	-----DGDIMDGMKDYK-----ANGPDKV 501
Db	710	TGQATTKPGQAGSALEGDSVQAQEQKQAPVPVPVPEAKAQVFTPPAPVNNKTENV 769
Qy	502	TMDGMNVYLTQYSDLNA-----KAVGVHYQFLDYNKPEVNDIPKGTISIEYADGKSV 556
Db	770	SKLD---YLEKLYEFLNTSYICHKYLIVSHSTWNEKILKQYKITKEBSKLSLCDPLDLL 826
Qy	557	FNIN-----DKRNNPGD---ETQOEHVINGKEYTSPFNDIKQIIDKTLNKIIV 603
Db	827	FNIQNNIPVWYSWFDLSNLSQLFMEIYEMVCNLYKLKDNDKIKNLLIEAKYKSTVS 886
Qy	604	KDFARNTTKPEFILNKDGEUSELXPHRVTTI QNGKEMSTIVSEBDFILPVYKGELEK 663
Db	887	KTLSSSMOPLSLTPODKPEVSANDDTGSHNTLNNLSKLFENILS-----LGNKN 937
Qy	664	GYOFDGEWISGPEGK-----KDAGVYVILSKDFTIKPVFKKIEEKEENKPTF--D 713
Db	938	IYQ-----ELIGQKSENFEKTLKDSDTFYNESFTNFVKSADDINSDESRRKLEED 993
Qy	714	VSKXKNQPNWHSQRKEDLQREBHSQKSOSTKDVATVLDDKNNISK--STNNP 771
Db	994	INKLKTLOLSFDLYNKKYKLERLFDKKTGVGYKMQIKKLTLLKEQLESKLNSLNP 1052

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2006, 22:35:57 ; Search time 52 Seconds
(without alignments)
1301.176 Million cell updates/sec

Title: US-10-067-385-8

Perfect score: 4026

Sequence: 1 KLGETAESKFNKLGNGKEGS.....ATVLDKNINSSKTTNNPNK 773

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 650591 segs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /EMC_Celerra_SIDS3/ptodata/2/iaa/5_COMB.pep:*
- 2: /EMC_Celerra_SIDS3/ptodata/2/iaa/6_COMB.pep:*
- 3: /EMC_Celerra_SIDS3/ptodata/2/iaa/7_COMB.pep:*
- 4: /EMC_Celerra_SIDS3/ptodata/2/iaa/H_COMB.pep:*
- 5: /EMC_Celerra_SIDS3/ptodata/2/iaa/PCTUS_COMB.pep:*
- 6: /EMC_Celerra_SIDS3/ptodata/2/iaa/RE_COMB.pep:*
- 7: /EMC_Celerra_SIDS3/ptodata/2/iaa/backfiles.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4026	100.0	773	2	US-09-590-991-8
2	3789.5	94.1	2138	2	US-09-583-110-5274
3	3048	75.7	637	2	US-09-107-433-3169
4	750.5	18.6	1529	2	US-09-107-433-4771
5	615	15.3	117	2	US-08-961-083-68
6	615	15.3	117	2	US-09-536-784-68
7	615	15.3	117	2	US-09-765-271-68
8	615	15.3	117	2	US-09-765-272A-68
9	218	5.4	1602	2	US-09-269-874A-7
10	218	5.4	1621	2	US-09-269-874A-3
11	218	5.4	1639	2	US-09-269-874A-5
12	203	5.0	1166	2	US-09-200-650E-7
13	200.5	5.0	1786	2	US-08-973-462-8
14	191.5	4.8	2504	2	US-09-328-352-5821
15	188.5	4.7	670	2	US-09-107-433-4976
16	188.5	4.7	708	2	US-09-583-110-3019
17	187.5	4.7	1529	1	US-08-728-470-10
18	187.5	4.7	1529	2	US-08-719-641-10
19	186.5	4.6	930	2	US-09-200-650E-3
20	186	4.6	1315	2	US-09-200-650E-5
21	185	4.6	2733	2	US-09-949-016-11433
22	185	4.6	3259	2	US-09-949-016-6507
23	184	4.6	1588	5	PCT-US93-07261-11
24	184	4.6	1663	5	PCT-US93-07261-16
25	181.5	4.5	778	2	US-09-134-001C-3868
26	179.5	4.5	746	2	US-09-710-279-652

27	179.5	4.5	1183	1	US-08-447-031A-2	Sequence 2, Appli
28	179	4.4	2142	2	US-09-540-236-3459	Sequence 3459, Ap
29	178	4.4	2375	2	US-09-538-092-1131	Sequence 1131, Ap
30	177.5	4.4	1600	1	US-08-617-697-10	Sequence 10, Appl
31	177	4.4	800	7	5183745-3	Patent No. 5183745
32	174.5	4.3	1073	2	US-09-206-942-49	Sequence 49, Appl
33	174.5	4.3	1073	2	US-10-193-764-45	Sequence 45, Appl
34	174.5	4.3	1079	2	US-09-206-942-47	Sequence 47, Appl
35	174.5	4.3	1079	2	US-10-193-764-43	Sequence 43, Appl
36	174.5	4.3	10182	2	US-09-134-001C-3159	Sequence 3159, Ap
37	174	4.3	3070	2	US-09-961-403-7	Sequence 7, Appli
38	173.5	4.3	2353	2	US-09-377-155-33	Sequence 33, Appl
39	173.5	4.3	2353	2	US-08-913-942-4	Sequence 4, Appli
40	173.5	4.3	2353	2	US-09-669-974-33	Sequence 33, Appl
41	173.5	4.3	2353	2	US-09-797-862-33	Sequence 33, Appl
42	173.5	4.3	2353	2	US-09-684-707-4	Sequence 4, Appli
43	173.5	4.3	2411	2	US-09-268-347-36	Sequence 36, Appl
44	173	4.3	930	2	US-09-134-001C-5314	Sequence 5314, Ap
45	173	4.3	930	2	US-09-386-962C-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1

US-09-590-991-8
; Sequence 8, Application US/09590991
; Patent No, 6887480
; GENERAL INFORMATION:
; APPLICANT: Adamou, John
; APPLICANT: Choi, Gil
; TITLE OF INVENTION: Streptococcus Pneumoniae Proteins and Vaccines
; FILE REFERENCE: 469201-475
; CURRENT APPLICATION NUMBER: US/09/590,991
; CURRENT FILING DATE: 2000-06-09
; EARLIER APPLICATION NUMBER: U.S. 60/138,453
; EARLIER FILING DATE: 1999-06-10
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 773
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-590-991-8

Query Match	100.0.0%;	Score	4026;	DB 2;	Length	773;			
Best Local Similarity	100.0.0%;	Pred. No.	3.6e-254;						
Matches	773;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	KLGEIAESKFNKLGNGKEGSLKDDTTGVVEHHHQQNEESIKESKSFITDRNISTIRDFENK	60						
Db	1	KLGEIAESKFNKLGNGKEGSLKDDTTGVVEHHHQQNEESIKESKSFITDRNISTIRDFENK	60						
Qy	61	DLKKLKKKPREVDFTSETGKMEEEYDYKYDDKGNIIAYDDGTDLEYETEKLDIEISKI	120						
Db	61	DLKKLKKKPREVDFTSETGKMEEEYDYKYDDKGNIIAYDDGTDLEYETEKLDIEISKI	120						
Qy	121	YGVLSPSKDGHPFELGKISNVSKNAKVYGGNNYKSIIEIKATKYDFHKTMTFDLYANIND	180						
Db	121	YGVLSPSKDGHPFELGKISNVSKNAKVYGGNNYKSIIEIKATKYDFHKTMTFDLYANIND	180						
Qy	181	IVDGLAFAGDMRLFKVNDNDQKAEIKIRPEKIKETKSEYPYVSSYGVNIELGEGDLSKN	240						
Db	181	IVDGLAFAGDMRLFKVNDNDQKAEIKIRPEKIKETKSEYPYVSSYGVNIELGEGDLSKN	240						
Qy	241	KPDNLTKMESGKIYSDSEKQQYLLKDNIIILRKGVALKVTTNPGKTDMLSGNGVYSKEDI	300						
Db	241	KPDNLTKMESGKIYSDSEKQQYLLKDNIIILRKGVALKVTTNPGKTDMLSGNGVYSKEDI	300						
Qy	301	AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFKMDKGEA	360						
Db	301	AKIQKANPNLRALSETTIYADSRNVEDGRSTQSVLMSALDGFNIIRYQVFTFKMDKGEA	360						

Qy 361 IDKGNLVTSSKLVLFKGDDKEYTGEDKFNVEAIKEDGSMFLFDTKPVNLSMDKNYFNP 420
Db 361 IDKGNLVTSSKLVLFKGDDKEYTGEDKFNVEAIKEDGSMFLFDTKPVNLSMDKNYFNP 420
Qy 421 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVDNYLIYGLDHDNTRDFNKLNVK 480
Db 421 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVDNYLIYGLDHDNTRDFNKLNVK 480
Qy 481 DGDIMDMGMDYKANGFPDVKVTDMDGNVYLQGTYSDLNAKAVGVHYQFLYDNVKNPEVNID 540
Db 481 DGDIMDMGMDYKANGFPDVKVTDMDGNVYLQGTYSDLNAKAVGVHYQFLYDNVKNPEVNID 540
Qy 541 PKGNTSIEYADGKSVWFNINDKRNNGPDGEIQEHIYINGKEYTSFNDIKQIIDKTLNIK 600
Db 541 PKGNTSIEYADGKSVWFNINDKRNNGPDGEIQEHIYINGKEYTSFNDIKQIIDKTLNIK 600
Qy 601 IVVKDFARNTTVKFEFLNKDTGEVSELKPHRVTVTIQNGKEMSTIYSEEDFILPVYKGE 660
Db 601 IVVKDFARNTTVKFEFLNKDTGEVSELKPHRVTVTIQNGKEMSTIYSEEDFILPVYKGE 660
Qy 661 LEKGQFDGWEISGFEGKDGAGYVNLKSDTFIKPVFKKIEEKEEENKPTFDVSKKDN 720
Db 661 LEKGQFDGWEISGFEGKDGAGYVNLKSDTFIKPVFKKIEEKEEENKPTFDVSKKDN 720
Qy 721 PQVNSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 773
Db 721 PQVNSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 773

RESULT 2
US-09-583-110-5274
; Sequence 5274, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
; FILE REFERENCE: PATH00-07A
; CURRENT FILING DATE: 2000-05-26
; PRIOR FILING DATE: 1997-07-02
; PRIOR FILING DATE: 1998-06-30
; PRIOR FILING DATE: 1998-05-12
; PRIOR FILING DATE: 1997-07-02
; NUMBER OF SEQ ID NOS: 5322
; SEQ ID NO 5274
; LENGTH: 2138
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-583-110-5274

Query Match 94.1%; Score 3789.5; DB 2; Length 2138;
Best Local Similarity 94.8%; Pred. No. 3.7e-238;
Matches 733; Conservative 15; Mismatches 24; Indels 1; Gaps 1;
Qy 1 KLGEIAESKFNGLNGKEGSLKDDTTGVEHHQNESESIEKSSFTIDRNIISTIRDFENK 60
Db 1333 KLGEIPESKFNGLNKKVDDSLNKETAVERNLLVDNQSIGKSLFNHHTIISTIRDFENK 1392
Qy 61 DLKXLIKKKREVDFTSETQKMEEDYKVDKGNIIAYDGDGTLEYETKLEIKSKI 120
Db 1393 DLKXLIKKKYQEDDFVN-GGTRTVERDYKDDKGNIIAYDGDGTLEYETKLEIKSKI 1451
Qy 121 YGVLSPSKDGHEILGKISNKNKAYVGNVYKSIETKATYDFHSTMTFIDLYANIND 180
Db 1452 YGVLSPSKDGHEILGKISNKNKAYVGNVYKSIETKATYDFHSTMTFIDLYANIND 1511
Qy 181 IVDGLAFAGDMRLFKVNDQKKAETKIRMPKIKETKSEYYPVSSYGNVIELGEGDLSKN 240
Db 1512 IVDGLAFAGDMRLFKVNDQKKAETKIRMPKIKETKSEYYPVSSYGNVIELGEGDLSKN 1571

Qy 241 KPNLTKMESGKIYSSEKQOYLLKNNIILRKGYALKVTTYNPGKTDMLBNGVYSKEDI 300
Db 1572 KPNLTKMESGKIYSSEKQOYLLKNNIILRKGYALKVTTYNPGKTDMLBNGVYSKEDI 1631
Qy 301 AKIQKANPNLRALESSTIYADSRNVEDGRSTQSVLMSALDGFNLIIRYQVTFKQNDKGEA 360
Db 1632 AKIQKANPNLRALESSTIYADSRNVEDGRSTQSVLMSALDGFNLIIRYQVTFKQNDKGEA 1691
Qy 361 IDKGNLVTSSKLVLFKGDDKEYTGEDKFNVEAIKEDGSMFLFDTKPVNLSMDKNYFNP 420
Db 1692 IDKGNLVTSSKLVLFKGDDKEYTGEDKFNVEAIKEDGSMFLFDTKPVNLSMDKNYFNP 1751
Qy 421 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVDNYLIYGLDHDNTRDFNKLNVK 480
Db 1752 SKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVDNYLIYGLDHDNTRDFNKLNVK 1811
Qy 481 DGDIMDMGMDYKANGFPDVKVTDMDGNVYLQGTYSDLNAKAVGVHYQFLYDNVKNPEVNID 540
Db 1812 DGDIMDMGMDYKANGFPDVKVTDMDGNVYLQGTYSDLNAKAVGVHYQFLYDNVKNPEVNID 1871
Qy 541 PKGNTSIEYADGKSVWFNINDKRNNGPDGEIQEHIYINGKEYTSFNDIKQIIDKTLNIK 600
Db 1872 PKGNTSIEYADGKSVWFNINDKRNNGPDGEIQEHIYINGKEYTSFNDIKQIIDKTLNIK 1931
Qy 601 IVVKDFARNTTVKFEFLNKDTGEVSELKPHRVTVTIQNGKEMSTIYSEEDFILPVYKGE 660
Db 1932 IVVKDFARNTTVKFEFLNKDTGEVSELKPHRVTVTIQNGKEMSTIYSEEDFILPVYKGE 1991
Qy 661 LEKGQFDGWEISGFEGKDGAGYVNLKSDTFIKPVFKKIEEKEEENKPTFDVSKKDN 720
Db 1992 LEKGQFDGWEISGFEGKDGAGYVNLKSDTFIKPVFKKIEEKEEENKPTFDVSKKDN 2051
Qy 721 PQVNSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 773
Db 2052 PQVNSQLNESHKEDLQREHSQKSDSTKDVATVLDKNNISSKSTTNNPNK 2104

RESULT 3
US-09-107-433-3169
; Sequence 3169, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007

```

;
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3169:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 637 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...637
; SEQUENCE DESCRIPTION: SEQ ID NO: 3169:
US-09-107-433-3169

Query Match 75.7%; Score 3048; DB 2; Length 637;
Best Local Similarity 99.3%; Pred. No. 1.5e-190;
Matches 585; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 185 LAFAGDMRLFVKDNDQKAEIKIRPEKIKETKSEYPYVSSYGNVIELGEGDLSKNKPDN 244
DB 15 LAFAGDMRLFVKDNDQKAEIKIRPEKIKETKSEYPYVSSYGNVIELGEGDLSKNKPDN 74
QY 245 LTKMESGKIYSSEKQOYLKKNIIIRKGYALKVTTYPNGKTDMLGNGVYSKEDIKIQ 304
DB 75 LTKMESGKIYSSEKQOYLKKNIIIRKGYALKVTTYPNGKTDMLGNGVYSKEDIKIQ 134
QY 305 KANPNLRALSETTIYADSRNVEDGSRSTQVLSALDGNIIIRYQVTFPMNDKGAIDKD 364
DB 135 KANPNLRALSETTIYADSRNVEDGSRSTQVLSALDGNIIIRYQVTFPMNDKGAIDKD 194
QY 365 GNLVTDSSKLVLFKGDDKEYTGEDKFNVBAIKEDGSMLEFIDTPVNLSDMKNYFNPSKSN 424
DB 195 GNLVTDSSKLVLFKGDDKEYTGEDKSNVEAIKEDGSMLEFIDTPVNLSDMKNYFNPSKSN 254
QY 425 KIYVRNPFFYLRGKISDKGFFNWLVRNESVVDNLYLYGDLHIDNTRDFNKLNVKGDGI 484
DB 255 KIYVRNPFFYLRGKISDKGFFNWLVRNESVVDNLYLYGDLHIDNTRDFNKLNVKGDGI 314
QY 485 MDWGMKDYKANGFPDKVTDMDGNVLOTGYSDLNAKAVGVHVOFLYDNVKNPEVNDPKGN 544
DB 315 MDWGMKDYKANGFPDKVTDMDGNVLOTGYSDLNAKAVGVHVOFLYDNVKNPEVNDPKGN 374
QY 545 TSIEYADGKSVVFNINDKRNNGFDEIOEQHIYINGKEYTSFNDIKQIIDKTLNKKIVVK 604
DB 375 TSIEYADGKSVVFNINDKRNNGFDEIOEQHIYINGKEYTSFNDIKQIIDKTLNKKIVVK 434
QY 605 DPARNTTVKGFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEDFILPVYKGELEKG 664
DB 435 DPARNTTVKGFILNKDTGEVSELKPHRVTVTIQNGKMSSTIVSEDFILPVYKGELEKG 494
QY 665 YQFDGWEISGPEGKDGAVINLSKDTFKPVFKKIEBKKEENKPTFDVSKKKNPOVN 724
DB 495 YQFDGWEISGPEGKDGAVINLSKDTFKPVFKKIEBKKEENKPTFDVSKKKNPOVN 554
QY 725 HSQLNESHKEDLQREHSQKSDSKTQVATVLDKNNISSKSTNNPNK 773
DB 555 HSQLNESHKEDLQREHSQKSDSKTQVATVLDKNNISSKSTNNPNK 603

RESULT 4
US-09-107-433-4771
; Sequence 4771, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGN
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

```

```

;
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4771:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1529 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...1529
; SEQUENCE DESCRIPTION: SEQ ID NO: 4771:
US-09-107-433-4771

Query Match 18.6%; Score 750.5; DB 2; Length 1529;
Best Local Similarity 79.7%; Pred. No. 2.6e-40;
Matches 149; Conservative 14; Mismatches 23; Indels 1; Gaps 1;

QY 1 KLGEIAESKFNKLGKSGSLKKTDTGVGHHQHNEBSIKKSSFTIDRNISTIRDFENK 60
DB 1336 KLGEIPESKFNKLGKSGSLKKTDTGVGHHQHNEBSIKKSSFTIDRNISTIRDFENK 1395
QY 61 DLKLLIKKKFREVDDFTSETGKRMEEYDYKYDDKGNIIAYDDGTDLLEYETEKLDIKSKI 120
DB 1396 DLKLLIKKKYQEDDFVN-GGTRTVRVDYKYDDKGNIIAYDDGTDLLEYETEKLDIKSKI 1454
QY 121 YGVLSPSKDGHEILGKISNVSKNAKYVGNYSIEIKATKYDPHSTMTFDFLYANIND 180
DB 1455 YGVLSPSKDGHEILGKISNVSKNAKYVGNYSIEIKATKYDPHSTMTFDFLYANIND 1514
QY 181 IVDGLAF 187
DB 1515 IVDGLISF 1521

RESULT 5
US-08-961-083-68
; Sequence 68, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville

```

```
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Dikette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961.083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-68

Query Match 15.3%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 5.5e-33;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 657 YKGELEKGYQFDGWEISGFEGKKGAGYVYNLSKDTFIKPVFKIEEKEENKPTFDVSK 716
Db 1 YKGELEKGYQFDGWEISGFEGKKGAGYVYNLSKDTFIKPVFKIEEKEENKPTFDVSK 60

Qy 717 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 773
Db 61 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 117

RESULT 6
US-09-536-784-68
; Sequence 68, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Dikette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961.083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; US-09-536-784-68
```

```
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-536-784-68

Query Match 15.3%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 5.5e-33;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 657 YKGELEKGYQFDGWEISGFEGKKGAGYVYNLSKDTFIKPVFKIEEKEENKPTFDVSK 716
Db 1 YKGELEKGYQFDGWEISGFEGKKGAGYVYNLSKDTFIKPVFKIEEKEENKPTFDVSK 60

Qy 717 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 773
Db 61 KKNPQVNHSQLNESHKEDLQREHSQKSDSTKDVTTATVLDKNNISSKSTTNNPNK 117

RESULT 7
US-09-765-271-68
; Sequence 68, Application US/09765271
; Patent No. 6887663
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Dikette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,271
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/536,784
; FILING DATE: <Unknown>
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-271-68
```

```
Query Match      15.3%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 5.5e-33;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YGGELEKGQFDGWEISGFEKKDAGYVNLKDTFIKPVFKKIEBKKEENKPTPDVSK 716
Db 1 YGGELEKGQFDGWEISGFEKKDAGYVNLKDTFIKPVFKKIEBKKEENKPTPDVSK 60

QY 717 KKDNPQVNSQLNESHKREDLQREHSQKSDSTKDVATATVLDKNNISSKSTNNPNK 773
Db 61 KKDNPQVNSQLNESHKREDLQREHSQKSDSTKDVATATVLDKNNISSKSTNNPNK 117

RESULT 8
US-09-765-272A-68
; Sequence 68, Application US/09765272A
; Patent No. 6929930
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and
; Vaccines
; NUMBER OF SEQUENCES: 454
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: Dell Latitude C610
; OPERATING SYSTEM: Windows 2000
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272A
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Lin J. Hymel
; REGISTRATION NUMBER: 45,414
; REFERENCE/DOCKET NUMBER: PB340P2C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 610-5790
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 68:
US-09-765-272A-68
```

```
Query Match      15.3%; Score 615; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 5.5e-33;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 657 YGGELEKGQFDGWEISGFEKKDAGYVNLKDTFIKPVFKKIEBKKEENKPTPDVSK 716
Db 1 YGGELEKGQFDGWEISGFEKKDAGYVNLKDTFIKPVFKKIEBKKEENKPTPDVSK 60

QY 717 KKDNPQVNSQLNESHKREDLQREHSQKSDSTKDVATATVLDKNNISSKSTNNPNK 773
Db 61 KKDNPQVNSQLNESHKREDLQREHSQKSDSTKDVATATVLDKNNISSKSTNNPNK 117
```

RESULT 9

```
US-09-269-874A-7
; Sequence 7, Application US/09269874A
; Patent No. 6933130
; GENERAL INFORMATION:
; APPLICANT: Bujard, Hermann
; TITLE OF INVENTION: Recombinant Process for Preparing a
; FILE OF INVENTION: Complete Malaria Antigen, GPI90/MSPI
; FILE REFERENCE: GRUE-003
; CURRENT APPLICATION NUMBER: US/09/269,874A
; PRIOR FILING DATE: 1999-08-02
; PRIOR APPLICATION NUMBER: PCT/EP97/05441
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: DE 19640817.2
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; US-09-269-874A-7
```

```
Query Match      5.4%; Score 218; DB 2; Length 1602;
Best Local Similarity 18.6%; Pred. No. 1.4e-05;
Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;

QY 23 KDTTG-VEHHHQENESIK-----EKSSFTIDRNIISTIRDFENKDKLKLKKKREVD 75
Db 213 KDNVGMEDYTKNNKKTIENTINELIEESKTKTDKNKNATKEE-----KKKLYQAOY 264

QY 76 FTSETGKEMEYDYKYDDKGMIIAYDDGTDLEYETEKLDEIKSKIYGVLSPSKDGHFEL 135
Db 265 DLSIYNKQLEE-----AHNLISV-----LE---KRIDTLK-----KNEIKELL 300

QY 136 GKISNVSNKAVYGNVNNYSIEIKATKYDFH-----SKMTFOLDYANINDIVDLAFA 188
Db 301 DKINEIKNPPPPANSNGTPTNLLDKNKKTEEHEKEIKETAKTIKFNIDSLFTDPL----- 354

QY 189 GDMRLFVKDNDQKAEIKIRMEPIKETKSEVP-----VVSXGNV-----IELGE-GDLSK 239
Db 355 -ELEYLLREKN-KNIDISAKVETKESTPEPNVGTYPPLSYNNALNELNSFGDLI- 411

QY 240 NKPDLNLTWESGKIYSDSEKQOYL--LKDNIL-----RKGYALKVTTVYNGKTDML 289
Db 412 -NPDYTKPEPSKIYTDNERKKFNEIKEIKIEKKIESDKKSYEDSKSLN-----DIT 466

QY 290 EG-----NGVYSKE-----DIAKTOKA-----NPNLRALSETTIYVADSRNVEDGRSTOSV 334
Db 467 KEYEKLLNEIYDSKFNNNIDLTNPEKMMGKRYSVKVEKLTHTHTPFASYENSKHLEKLT 526

QY 335 LMSALDGF--NII-----RYQVFTFMNDKGEAIDKGNLVTDSKLVLFKGDDEY 385
Db 527 ALKYMEDYSLRNI VVEKELKYKNLISKIENEIETLVE--NIKKDEEQLF-----EKKIT 579

QY 386 GEDKFNVEAIAEKDGSMLFIDTKPVNLSMDKNVFNPSKSKIYVVRNPFVLRGKISDKG 445
Db 580 KDNKPKDEKILEVSDIVKQVQKVL---MKNIDELKKTQLILKNVE--LKHNIHVPSY 634

QY 446 NWELRVNESVVDNYLIYGDHLIDNTRDNILK-----NVK----- 480
Db 635 KQENKQE-----PYLIIVLKKSIDKLKVPMPKVESLINEEKKNIKTEGQSDNSEPTEGEI 690

QY 481 -----DGDINDWGMKDYK----- 501
Db 691 TQQTATTPGQAGSALGDSVQAQEQKQAPVPVPEAKAQVPTTPAPVNNKTEV 750

QY 502 TDMGNYVLOTGYSDLNA-----KAVGVHYOFLYDYNKPEVNIIDPKGNTSIEYADKGVV 556
Db 751 SKLD---YLEKLYEFLNTSYICHKYLIVSHSTMNEKILKQYKIKTEESKSLSCDPLDL 807

QY 557 FNIN-----DKRNNPGDG---EIQEOHIYINGKEYTSFNIDIKOIIDKTLNIIKW 603
Db 808 FNIQNNIPVMYSMFDLSNLSQLFMEIYEMVCNLYKLKDNDKIKNLLSEAKKVS 867
```

```
Qy 604 KDFARNTTVKEFILNKDTGTVSELKPHRVTVTTIQNGKEMSTIVSEEDFILPVYKGELEK 663
Db 868 KTLSSSSMQPLSLTPQDKPEVSANDTSHSTNLNLSKLFENILS-----LGNKN 918
Qy 664 GYQFDGWEISGFEK-----KDGAVINLSKDTFIKPVFKKIEBKKEENKPTF--D 713
Db 919 IYQ-----ELIGQKSENFEYKILKDSDTFYNESFTNFVKSKADDINSLNDESKRKKLEED 974
Qy 714 VSKKDNQPNVHNSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSK-STTNNP 771
Db 975 INKLKTLQLSFDLYNKYKLERLFDKKTGVGYKQKIKLTLLEKQESKLSLNNP 1033

RESULT 10
US-09-269-874A-5
; Sequence 5, Application US/09269874A
; Patent No. 6933130
; GENERAL INFORMATION:
; APPLICANT: Bujard, Hermann
; TITLE OF INVENTION: Recombinant Process for Preparing a
; TITLE OF INVENTION: Complete Malaria Antigen, GP190/MSp1
; FILE REFERENCE: GRUE-003
; CURRENT APPLICATION NUMBER: US/09/269, 874A
; PRIOR FILING DATE: 1999-08-02
; PRIOR FILING DATE: 1997-10-02
; PRIOR FILING DATE: 1996-10-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1621
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-09-269-874A-5

Query Match 5.4%; Score 218; DB 2; Length 1621;
Best Local Similarity 18.6%; Pred. No. 1.4e-05;
Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;

Qy 23 KDTTG-VEHHHQENESIK-----EKSFTIDRNISTIRDENKDLKLIKKEFVDD 75
Db 232 KDNVGMEDYIKKKNKTIENTINELIESKTTIDKNKNATKEE-----KKLYQAY 283
Qy 76 FTSETGKRMEYDYKYDDKGNIIAYDDGTULEYETEKLEIKSIYGVLSPKDGHPFEL 135
Db 284 DLSIYNKQLEE-----AHLNLSV-----LE---KRIDLK-----KNENIKELL 319
Qy 136 GKISNVSKNAKYVYGNKYSEIKATKYDFH-----SKMTTFDLYANINDIVDGLAFA 188
Db 320 DKINEIKNPPANGSGTNPNTLLDKNKKIEHEKEIKIAKTIFKNIDSLFTDPL----- 373
Qy 189 GDMRLFVKNDQKAEIKIRMPKIKETKSEYP-----YVSSYGNV-----TELGE-GDLSK 239
Db 374 -ELEYLRKKN-KNIDISAKVETKESTPNPNGVTPYPLSYNDINNALNELNLSFGDLI- 430
Qy 240 NKPNLTLMESGKIYSDSEKQOYL--LKDNIL-----RKGVALKVTTYPNGKTDML 289
Db 431 -NPPDYTKPSKNITVDNERKKFNEIKIEKKIESDKSYEDRSKSLN---DIT 485
Qy 290 EG-----NGVYSKE-----DIAKTQKA-----NPNRLALSTETTYADSRNVEDGRSTQSV 334
Db 486 KEYEKLNLNIIYDSKFNNNIDITNPFQWGMGRKYSYKVEKLTTHNTFASYSKNHLEKLTK 545
Qy 335 LMSALDGE---NII-----RYQVTFQNDKGAIKDGKLVNLTSSKLVLFQKDDKEYT 385
Db 546 ALKTMEDYSLNIIIVVEKELKYKNLSKIENIEITLVE--NIKDEBELF-----EKKIT 598
Qy 386 GEDKFNVBALIKEDGSMFLFDTKPVNLSMDKNYFNPSKSNKIYVRNPFYLGKISDKGGF 445
Db 599 KDNKPKDEKILEVSDIVKVQVQKVL---MNKIDELKKTQLILKNVE--LKHNIHVPSNY 653
```

```
Qy 446 NMELRVNESVDNYLIYGDHLHDNTDEFNKL-----NVK----- 480
Db 654 KOENKQE-----PYLLIVLKKEDIKLHVMPKVESLINEKKNIKTSGQSDNSPESTGEI 709
Qy 481 -----DGDIMDMGMFDYK-----ANGFPDKV 501
Db 710 TGOATTTKPGQAGSALEGDSVQAQAEQKQAOQPPVPVPEAKAQQVPTPPAPVNNKTENV 769
Qy 502 TMDGNVYIQTGYSDLNA-----KAVGHYQFYLDNVKPEVNTDPKGNITSIEYADGKSVV 556
Db 770 SKLD---YLEKLYEFLNTSYICHKYLIVSHSTNNEKILKQYKITKEESKLSKCDPLDLL 826
Qy 557 FNIN-----DKRNNGFDG---EIQEOHIYINGKEYTSENDIKIIDLKTLNLIKIV 603
Db 827 FNIONNIPVMYSMFDLSNLSOLFMEIYEKEMVCNLYKLDKNDKIKNLLLEEAKKVTSTV 886
Qy 604 KDFARNTTVKEFILNKDTGTVSELKPHRVTVTTIQNGKEMSTIVSEEDFILPVYKGELEK 663
Db 887 KTLSSSSMQPLSLTPQDKPEVSANDTSHSTNLNLSKLFENILS-----LGNKN 937
Qy 664 GYQFDGWEISGFEK-----KDGAVINLSKDTFIKPVFKKIEBKKEENKPTF--D 713
Db 938 IYQ-----ELIGQKSENFEYKILKDSDTFYNESFTNFVKSKADDINSLNDESKRKKLEED 993
Qy 714 VSKKDNQPNVHNSQLNESHKREDLQREHSQKSDSTKDVATVLDKNNISSK-STTNNP 771
Db 994 INKLKTLQLSFDLYNKYKLERLFDKKTGVGYKQKIKLTLLEKQESKLSLNNP 1052

RESULT 11
US-09-269-874A-3
; Sequence 3, Application US/09269874A
; Patent No. 6933130
; GENERAL INFORMATION:
; APPLICANT: Bujard, Hermann
; TITLE OF INVENTION: Recombinant Process for Preparing a
; TITLE OF INVENTION: Complete Malaria Antigen, GP190/MSp1
; FILE REFERENCE: GRUE-003
; CURRENT APPLICATION NUMBER: US/09/269, 874A
; CURRENT FILING DATE: 1999-08-02
; PRIOR APPLICATION NUMBER: PCT/EP97/05441
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: DE 19640817.2
; PRIOR FILING DATE: 1996-10-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1639
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-09-269-874A-3

Query Match 5.4%; Score 218; DB 2; Length 1639;
Best Local Similarity 18.6%; Pred. No. 1.4e-05;
Matches 167; Conservative 164; Mismatches 340; Indels 228; Gaps 39;

Qy 23 KDTTG-VEHHHQENESIK-----EKSFTIDRNISTIRDENKDLKLIKKEFVDD 75
Db 232 KDNVGMEDYIKKKNKTIENTINELIESKTTIDKNKNATKEE-----KKLYQAY 283
Qy 76 FTSETGKRMEYDYKYDDKGNIIAYDDGTULEYETEKLEIKSIYGVLSPKDGHPFEL 135
Db 284 DLSIYNKQLEE-----AHLNLSV-----LE---KRIDLK-----KNENIKELL 319
Qy 136 GKISNVSKNAKYVYGNKYSEIKATKYDFH-----SKMTTFDLYANINDIVDGLAFA 188
Db 320 DKINEIKNPPANGSGTNPNTLLDKNKKIEHEKEIKIAKTIFKNIDSLFTDPL----- 373
Qy 189 GDMRLFVKNDQKAEIKIRMPKIKETKSEYP-----YVSSYGNV-----TELGE-GDLSK 239
Db 374 -ELEYLRKKN-KNIDISAKVETKESTPNPNGVTPYPLSYNDINNALNELNLSFGDLI- 430
Qy 240 NKPNLTLMESGKIYSDSEKQOYL--LKDNIL-----RKGVALKVTTYPNGKTDML 289
```

Db 431 -NPFDTKPSKNIYTDNERKFEINKEIKIEKKIBSDKSYEDRSKSLN-----DIT 485
Qy 290 EG-----NGVYSKE-----DIAIKQA-----NPNLRALSETTIYADSRNVEDGRSTQSV 334
Db 486 KEYEKLNLNEYDSKFNNIDLTFEKOOMGRYSYKVEKLTHTHTTFASYENSXHNLEKLTK 545
Qy 335 LMSALDGF---NI-----RYQVFTKMDKGEAIDKDNLTVDSSKLVLFGKDKKEYT 385
Db 546 ALKYMEDYSLRNIVWEKELKYKLNLSKIENBETLVE---NIKKDEQLF-----EKKIT 598
Qy 386 GEDKFNVEAIKEDGSMFLFTDTPVNLSDMKQVFNPSKSNKIYVRNPFYLRKISDKGKF 445
Db 599 KOENKPEKILEVSDIVKVQVQKVL---MNKIDELKTLQLILKNVE---LKHNIHVPNSY 653
Qy 446 NWELRVNESVVDNLYLGLHDINTREDNKL-----NVK----- 480
Db 654 KQENKQE-----PYLLVLKKEIDKLKVFMPKVESLINEEKKNIKTEGQSDNSEPSTGEI 709
Qy 481 -----DGDIMDMGMKDYK-----ANGFPDKV 501
Db 710 TGQATTKPQOAGSALEGDSVQAQAQEQQAQPPVPPVPEAKAQVPTPPAPVNNKTENV 769
Qy 502 TMDGNVYLQTVSDINA-----KAVGVHYQFLYDNVKNPEVNIDPKGNTSIYADGKSVV 556
Db 770 SKLD---YLEKLYEFLNTSYICHKYLVSHTMNEKILKQYKTKTEESKLSGCDPLDL 826
Qy 557 FNIN-----DKRNNFGD---EIOEQHIYINGKEYTSFNDIKQIIDTLAIKIVV 603
Db 827 FNIOQNPVWYMFDSNLNSLSQLFMEIYEKEMVNCNLYKLKDNKIKNLLEAAKVSTSV 886
Qy 604 KDFARNTTVKEFILNKDTEGEVSELKPHRVTVTIQNGKEMSTIVSEEDFILPVYKGELEK 663
Db 887 KTLSSSSMQPLSITPDQKEVSANDDTSHSTWNLNSLKLFPENLS-----LGKXKN 937
Qy 664 GYQDFGWEISGPEGK-----KDAGYVINLSKDTFFIKVPKPKIEEKEEENKPTF--D 713
Db 938 IQY-----ELIGQKSENFEYKILKDSDFYNESFTNFVKSADDINSLNDESKRKKLEED 993
Qy 714 VSKKXNDPQVNHSQLNESHKEDLQREHSQKSDSTKQVATVLDKNISSK-STNNP 771
Db 994 INKLKTLQLSPDLNKNYKKLKLERLDFDKKTVGKYKMQIKKLTLLKEQLSKLSLNNP 1052

RESULT 12

US-09-200-650E-7
; Sequence 7, Application US/09200650E
; Patent No. 6680195
; GENERAL INFORMATION:
; APPLICANT: Patti, Joseph M.
; APPLICANT: Foster, Timothy J.
; APPLICANT: Hook, Magnus A.O.
; APPLICANT: Perkins, Samuel L.
; APPLICANT: Deirdre Ni
; APPLICANT: Perkins, Samuel L.
; TITLE OF INVENTION: Extracellular Matrix-Binding Proteins from Staphylococcus aureus
; FILE REFERENCE: P06283US2/BAS
; CURRENT APPLICATION NUMBER: US/09/200,650E
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,815
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: 60/098,427
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1166
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-200-650E-7

Query Match 5.0%; Score 203; DB 2; Length 1166;
Best Local Similarity 21.6%; Pred. No. 8.7e-05;
Matches 184; Conservative 120; Mismatches 291; Indels 258; Gaps 53;

Qy 33 QENESSIKE--KSSFTIDENISTIRDFENKDLKLLKKKFR-----EYDDFTS 78
Db 226 KNNPEKLKELVNDNONTDSTKPTAPTSVAPKRVNAKQRFAPAQAASVANNVNDLTK 285
Qy 79 ETGKEMEYDYKDDKGNIIAYDDGTDLLEYETE-----KLDIEKSKIYGLVS 125
Db 286 VTKQITIK-----VGDKDNVAAAHDKDIEYDTEFTIDNKKVKGDTWNTYDNVIPSJLT 341
Qy 126 PSKD-----GHEIILKISNVSNAK---VYGNVYKSIIEIKATKYDFHISK----- 168
Db 342 DKNDPIDITDPSGEVIAGTFDKATKQITPTFDYVDKYEDIKSRLLTLYSYDKKTVPNE 401
Qy 169 ---TWTF-----DLVANIN-DIVDGLAFAGD---MRLFKV-DNDQKKAIEKIRM-PEKIK 214
Db 402 TSLNLTPTATAGKETSONVTVDYQDPWH-GDSNIOISIFTKLDEDKQTTEQQIYVNPPLKKS 460
Qy 215 ETKSY-----PVSSYGNVIELGEGD--LSKNKPNLTKM-----BSGKIYSSEKQ 261
Db 461 ATNTKVDIAGSQVDDYGN-IKLGNGSTIIDQNTETKVKVKNVSDQQLPQSNRIYDFSQYED 519
Qy 262 YLLK-DNIIILKGYALKVTTYNPGKTDMLBGNVGVSKEDIAKIQANPNLRALSETTIYA 320
Db 520 VTSQFDN---KKSFNNVATLDFG-----DINSAYI-----IKVSVKTYPTS 558
Qy 321 DSR-NVEDGRSTQSVLMSALDGFNIIRYOVTFFKNDKG-----E 359
Db 559 DGLDIAOGTSNRIT--DKYGYNYAGYSNFIIVTSNDTGGGDTVKPBEKLYKIGDYVWE 616
Qy 360 AIDKGNLVTDSKLVLFKDKDEYTBGDKFNVEAIKEDGSMFLFTDTPVNLSDMKNY-F 418
Db 617 DVDKDG-----VQDTSKEKPMANVL-VTLTYPDGT-----TKSVRTDANGHYEF 660
Qy 419 NPSKSNKIYV---RNPEFYLRGKIS-----DKGFNWEIWRVNSVVDNLYLIGDLHID 468
Db 661 GGLKDGETYTVKFETPTGTLTKVNGTTDGEKDSNGSSVTVKINGK-----DMSLD 712
Qy 469 -----NTRDFNKLNVKDG--DMDWGMKQYKANGFPDPKVTMDGNVYVLTQYSDL 517
Db 713 TGFYKPKYNGYQVYVWEDTNKDGIDQANEPGDKVKVT-----LKDSTGKV-IGTTTDA 766
Qy 518 NAKAVGVHYQFLYDNVKNPEVNIDPKGNTSIY---ADGKSVVFN--INDKRNNGFDEIQ 572
Db 767 SGK-----YKF-----TDLD-NGNVTVEFEIPAGYTPVKNTTADDKOSNGL----- 807
Qy 573 EOHYINGKEYTSFNDIKQIIDKTLNLIKIVVKDFARNT--TVKEFI---LNKDTGEVSEL 627
Db 808 -----TTTGVIKDADNMTLD-----RGFYKPKYSLGYVYVYDSNKKDGKQDSTE 851
Qy 628 KPHR-VTVTIQNGK-EMSSSTIVSEEDFILPVYKGELEKGYQDFGWEISGFEGKQDAGYVI 685
Db 852 KGIKDVTVLQNEKEVIGITKTIDEN-----GK-----YRPD----- 883
Qy 686 NLSKDTFIKPVFKKIEEKEEENKPTFDVSKKDNQPV-----NHSQ--LNESHKEDLIQ 738
Db 884 NLDGSKY-KVIFEXFAGITQIVTWTED-DKADGGEVDVITDHDHDFDLNGYFEEDTS 941
Qy 739 REHSQKSDSTKD 751
Db 942 DSDSDSDSDSDSD 954

RESULT 13

US-08-973-462-8
; Sequence 8, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06

THIS PAGE BLANK (uapfo)